IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

*Digital Object Identifier 10.1109/ACCESS.2019.DOI*

# Enabling Reproducible Research in Sensor-Based Transportation Mode Recognition with the Sussex-Huawei Dataset

**LIN WANG[1,3], HRISTIJAN GJORESKI[1,4], MATHIAS CILIBERTO[1], SAMI MEKKI[2], STEFAN VALENTIN[2,5] (Member, IEEE) and DANIEL ROGGEN[1] (Member, IEEE)**

[1]Wearable Technologies Laboratory, Sensor Technology Research Centre, University of Sussex, Brighton BN1 9QT, U.K. (e-mail: {w23, h.gjoreski, m.ciliberto}@sussex.ac.uk; daniel.roggen@ieee.org)
[2]Mathematical and Algorithmic Sciences Lab, PRC, Huawei Technologies France, 92100 Boulogne-Billancourt, France (e-mail: sami.mekki@huawei.com)
[3]Centre for Intelligent Sensing, Queen Mary University of London, London E1 4NS, U.K. (e-mail: lin.wang@qmul.ac.uk)
[4]Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius University, Skopje 1000, Macedonia (e-mail: hristijang@feit.ukim.edu.mk)
[5]Department of Computer Science, Darmstadt University of Applied Sciences, Darmstadt 64295, Germany (e-mail: stefan.valentin@h-da.de)

Corresponding author: Lin Wang (e-mail: w23@sussex.ac.uk).

This work was supported by Huawei Technologies within the project "Activity Sensing Technologies for Mobile Users".

**ABSTRACT** Transportation and locomotion mode recognition from multimodal smartphone sensors is useful to provide just-in-time context-aware assistance. However, the field is currently held back by the lack of standardized datasets, recognition tasks and evaluation criteria. Currently, recognition methods are often tested on ad-hoc datasets acquired for one-off recognition problems and with differing choices of sensors. This prevents a systematic comparative evaluation of methods within and across research groups. Our goal is to address these issues by: i) introducing a publicly available, large-scale dataset for transportation and locomotion mode recognition from multimodal smartphone sensors; ii) suggesting twelve reference recognition scenarios, which are a superset of the tasks we identified in related work; iii) suggesting relevant combinations of sensors to use based on energy considerations among accelerometer, gyroscope, magnetometer and GPS sensors; iv) defining precise evaluation criteria, including training and testing sets, evaluation measures, and user-independent and sensor-placement independent evaluations. Based on this, we report a systematic study of the relevance of statistical and frequency features based on information theoretical criteria to inform recognition systems. We then systematically report the reference performance obtained on all the identified recognition scenarios using a machine-learning recognition pipeline. The extent of this analysis and the clear definition of the recognition tasks enable future researchers to evaluate their own methods in a comparable manner, thus contributing to further advances in the field. The dataset and the code are available online[a].

[a]http://www.shl-dataset.org/

**INDEX TERMS** activity recognition, feature selection, mobile sensing, multimodal sensor fusion, reference dataset, transportation mode recognition

## I. INTRODUCTION

Today's mobile phones come equipped with a rich set of sensors, including accelerometer, gyroscope, magnetometer, global positioning system (GPS) and others, which can be used to discover user activities and context [1], [2]. Transportation and locomotion modes are an important element of the user's context that denotes how users move about, such as by walking, running, cycling, driving car, taking bus or subway (Fig. 1) [3], [4]. Transportation and locomotion mode recognition is useful for a variety of applications, such as human-centered activity monitoring [5], [6], individual environmental impact monitoring [7], [8],

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



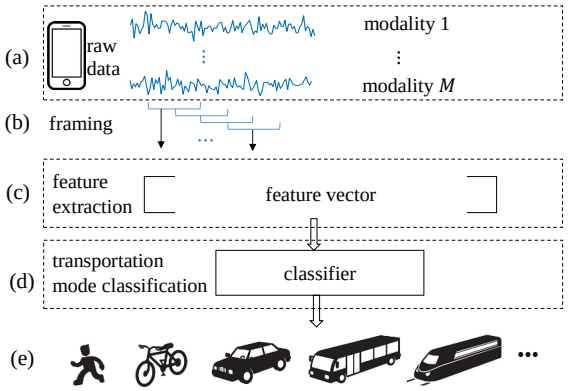


**FIGURE 1.** Transportation mode recognition from mobile phone sensor data is generally addressed using streaming machine learning techniques. The data from multimodal sensors (a) are segmented into short frames of sensor signals (b) on which features are computed yielding a feature vector (c). A classifier (d) then maps the feature vector in one of the transportation classes (e).

just-in-time distributed intelligent service adaptation [9], [10], and implicit human computer interaction [11]–[13].

In recent years, there have been numerous studies showing how to recognize transportation modes from multimodal smartphone sensor data with machine learning techniques [3], [4], [14]. However, there is still not a well-recognized dataset that can be used for performance evaluation by the research community. To date, most research groups assess the performance of their algorithms using their own collected data, which cover a different number of transportation activities and sensor modalities. Due to the complexity of the data collection procedure and the need to protect participant privacy, these ad-hoc datasets often have a short duration and remain private. This prevents the comparison of different approaches in a replicable and fair manner within and across research groups and impedes the progress in this research area.

Considering this we believe that there is a need for advancing reproducible research in sensor-based transportation and locomotion mode recognition. This requires publicly available datasets, common recognition tasks (i.e. number and type of transportation and locomotion classes to recognize), common combinations of sensors to use, and identical evaluation procedures. Ideally, these datasets should contain sufficient transportation activities, sensor modalities, and recording duration in order to verify the versatility of the developed algorithms. The recognition tasks and evaluation measures should cover the most common application needs currently identified by the research community and should be forward looking to accommodate upcoming application needs. The objective of this paper is to support reproducible and comparable research within and across research groups in the field of transportation mode recognition.

Other research communities have acknowledged the need to establish reference recognition tasks to support scientific advances in their field. This is the case, for example, in computer vision with the PASCAL Visual Object Classes challenge [15] or the ImageNet Large Scale Visual Recognition Challenge [16] and in speech recognition with the CHiME corpus and recognition challenge [17].

We have previously introduced the large-scale Sussex-Huawei Locomotion (SHL) dataset which was recorded over a period of seven months by three participants engaging in eight transportation activities in real-life setting, including Still, Walk, Run, Bike, Car, Bus, Train and Subway [18], [19]. The dataset contains multimodal data from 16 smartphone sensors, which are precisely annotated and amount up to 2800 hours. We use this dataset as a baseline to establish a standardized evaluation framework and to promote reproducible research in the field. The contribution of the paper is summarized as below.

1) Survey of the state-of-the-art. We conducted a comprehensive literature review over the 30 academic articles published in recent years on the problem of transportation mode recognition. We conducted a very thorough state-of-the-art analysis in terms of dataset availability, including sensor modalities and number of classes, and in terms of recognition pipeline characteristics, including processing window size, used features and classifiers, postprocessing techniques. To our knowledge, this is one of the most comprehensive literature reviews in the field of transportation mode recognition from mobile devices. This will give readers a clear understanding of the state-of-the-art in this field. Through state-of-the-art analysis, we found out that the lack of standard dataset, unified recognition task and evaluation criteria prevents a fair comparison between different research groups, and thus holds back the progress of research in the field. This paper thus aims to address these challenges with the SHL dataset, which is one of biggest and publicly available dataset in the field.

2) Standardized evaluation framework with baseline implementation. To enable reproducible research, we precisely defined standardized evaluation process. This academic contribution will enable researchers to compare methods "likes to likes": they will be able to use the exact same tasks to compare methods, therefore helping to clearly identify benefits of novel methods. The framework consists of 12 evaluation scenarios, 6 groups of sensor modalities, and 3 types of cross-validation schemes, leading to 729 recognition tasks in total. These tasks are defined considering both the sensor modalities of the SHL dataset and the various recognition tasks we identified from our related work review. We implemented a basic recognition pipeline to report baseline performance for all these tasks and will make the source code publicly available. Researchers in this field will have several options to develop new methods based on our evaluation framework. They will be able i) to evaluate their new newly develop algorithms with this dataset and the evaluation tasks; ii) to apply the baseline recognition system with their own dataset; iii) to create the recognition tasks based on the recommendation of the paper with their own dataset and own algorithms and compare with the baseline results reported in the paper. We believe this will advance the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



progress the research in this field significantly.

3) Feature analysis and feature selection based on the SHL dataset. The large amount of data in the dataset allows us conduct a thorough analysis to investigate the ability of a large set of features to distinguish between any two transportation activities. We proposed a large set of features (2727 in total), which include all the features considered in the literature plus additional features computed based on the time-domain quantile values and frequency-domain subband energies. We proposed a feature analysis method based on mutual information. The method visualizes the ability of each feature and sensor modality to distinguish any two transportation activities. We further proposed a feature selection method based on pair-wise maximum-relevance-minimum-redundency (MRMR) which selects a small set of features that are suitable for recognizing the 8 class activities. The large set of features, the feature analysis and visualization, and the feature selection method are new in this research field. This will give readers a better understanding of the dataset, and will help them to identify better features and develop new recognition methodologies in their work. Thanks to this, our work is the first to show clearly which frequency band contains the most valuable information to distinguish transportation modes, and it is the first to clearly identify that magnetic field sensors provides additional critical information to distinguish between modes of transport, contrarily to a common held assumption.

The organization of the paper is as follows. After reviewing the state of the art in Sec. II, we introduce the SHL dataset in Sec. III and recommend a list of standard transportation mode evaluation tasks in Sec. IV. We perform feature analysis in Sec. V and establish the baseline performance in Sec. VI. After discussions in Sec VII we draw conclusions in Sec. VIII.

## II. STATE OF THE ART

### A. APPROACHES TO TRANSPORTATION MODE RECOGNITION

Fig. 1 depicts a basic processing pipeline for predicting the transportation mode using the multimodal sensors embedded in the smartphone carried by the user. The multimodal sensor data (such as inertial and GPS) are first segmented into frames with a sliding window. The data in each frame is used to compute a vector of features. These feature vectors are processed by a classifier which aims to recognize the transportation mode of the user.

Table 1 gives a comprehensive summary on the literatures that work on transportation and locomotion mode recognition, which can be categorized into three families: inertial based, location based, and hybrid. *Inertial based* approaches employ inertial sensors to detect the acceleration (accelerometer), rotation (gyroscope) and ambient magnetic field (magnetometer) of the mobile device, and predict the transportation mode of the user based on the motion pattern of the mobile device itself [20]–[35], [54]. *Location based* approaches employ the GPS receiver to detect the location of the mobile device, and predict the transportation mode based on the motion pattern of the user, such as GPS speed, GPS acceleration, and the trajectory of the trip [38]–[47]. Geographic information system (GIS) can be used to further improve the recognition accuracy by exploiting information such as the closeness to train stations, bus stops, rail lines, and roads [43], [44], [46]. *Hybrid* approaches combine inertial and GPS sensors to predict the transportation mode and thus usually perform better than using one modality alone [48]–[53]. We analyze the state of the art from four aspects: dataset and sensor modality, type of classifier, decision window, and number of classes.

*Dataset and modality*. Due to costs and time required to collect and annotate datasets, most research groups working with inertial sensors used datasets with limited duration (dozens of hours). Due to the earlier availability of accelerometers on mobile phones, the majority of datasets to date include accelerometers as the sole modality. Some exceptions include three datasets with multiple modalities (accelerometer, gyroscope, magnetometer) but a limited duration of 12 hours [24], 25 hours [23], and 13 hours [55], respectively; two datasets with single modality (accelerometer) but a long duration of 100 hours [25] and 890 hours [29], respectively; and a large dataset with multiple inertial modalities and a long duration of 8311 hours [20], [21]. A common problem is that none of dataset mentioned above is publicly available except [55] with 13 hours of data. Most research groups working with GPS sensors only used large dataset containing hundreds to thousands of trips. Geolife, a large dataset with GPS information from 9043 trips is publicly available [56].

There are only a few research groups working on hybrid approachers, including [49], [50], [52] who used datasets with a duration between 100 to 350 hours. Currently all the datasets reported with hybrid approaches contain only two modalities, i.e. GPS and accelerometer. All these datasets have much less modalities than the SHL dataset. The richest dataset [50] contain 2 modalities and 355 hours of data, which is significantly less than SHL with 16 modalities and 2800 hours of data.

*Number of classes*. Most papers reviewed report a different classification task, ranging from recognizing three transportation classes (e.g. Walk, Car and Train [34]) to ten (e.g. Still, Walk, Run, Bike, Motorcycle, Car, Bus, Subway, Train, and High speed rail [20]). Among various transportation activities, the most frequently considered ones are Still, Walk, Run, Bike, Car, Bus, Train and Subway. The variety of transportation mode recognition tasks creates a problem of reproducible research.

*Decision window size*. The sensor data are divided into frames with a sliding window and processed per frame. There is a trade-off when choosing the size of the sliding window, which affects the classification accuracy, response time (latency), and memory size [22], [26]. The preferred choice of window size varies in the papers we reviewed. Generally, inertial based approaches use a short window size

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



**TABLE 1.** Approaches for transportation mode recognition using inertial ($\mathcal{I}$), location ($\mathcal{L}$) and inertial-location hybrid ($\mathcal{H}$) sensors. Key: Acc - Accelerometer; Gyr - Gyroscope; Mag - Magnetometer; Bar- Barometer; Mic - Microphone.

| Approach | Reference | Dataset | | | Transportation classes | Classifier | Window |
|---|---|---|---|---|---|---|---|
| | | Availability | Duration | Modality | | | |
| $\mathcal{I}$ | [20], [21] | Private | 8311 h | Acc, Mag, Gyr | Still, Walk, Run, Bike, Motorcycle, Car, Bus, Subway, Train, HSR | DT, KNN, SVM, DNN | 17.2 s |
| | [22] | | 8311 h | Acc, Mag, Gyr | Still, Walk, Run, Bike, Motorcycle, Car, Bus, Subway, Train, HSR | DT, AdaBoost, SVM | 17.2 s |
| | [23] | | 12 h | Acc, Mag, Gyr | Still, Walk, Bike, Bus, Car, Subway | DT, KNN, SVM | 8 s |
| | [24] | | 25 h | Acc, Mag, Gyr | Walk, Run, Bike, Bus, Car | KNN, SVM, DT, Bagging, RF | 1 s |
| | [25] | | 150 h | Acc | Still, Walk, Bus, Car, Train, Subway, Tram | Adaboost+HMM | 1.2 s |
| | [26] | | 3 h | Acc, Gyr, Mag, Bar | Walk, Run, Bike, Bus, Car, Subway | SVM | 12.8 s |
| | [27] | | 4 h | Acc | Still, Walk, Run, Bike, Car | KNN, QDA | 7.5 s |
| | [28] | | 30 h | Acc | Walk, Bike, Bus, Subway, Car, Drive | DT | 8 s |
| | [29] | | 890 h | Acc | Walk, Bike, Car, Train | SVM, Adaboost, DT, RF | 7.8 s |
| | [30] | | NA | Acc | Walk, Bus, Car, Train | Thresholding | 5 s |
| | [31] | | 8.9 h | Acc | Walk, Run, Bike, Bus, Car, Train | SVM | 5 s |
| | [32] | | 29 h | Acc | Still, Walk, Bike, Bus, Car, Train, Tram, Subway, Boat, Plane | DT, RF, BN, NB | 5 s |
| | [33] | | 2.5 h | Acc | Sit, Stand, Walk, Run, Bike, Car | DT, NB, kNN, SVM | NA |
| | [34] | | 9 h | Acc | Walk, Car, Train | NB, SVM | 4 s |
| | [35] | | 3 h | Acc, Gyr, Mag | Walk, Run, Bike, Car | kNN, DT, RF | 5 s |
| | [36] | | 20 h | Acc | Still, Walk, Bike, Bus, Car, Train, Subway, Motorcycle | DT | 10 s |
| | [37] | | 47 h | Bar | Still, Walk, Vehicles | Thresholding | 200 s |
| $\mathcal{L}$ | [38], [39] | Public (Geolife) | 7112 trips | GPS | Walk, Bike, Bus, Car | DT, SVM, BN, CRF, Graph | whole trip |
| | [40] | | 17621 trips | GPS | Walk, Bike, Bus and taxi, Car, Train, Subway | kNN, DT, SVM, RF, XGBoost, GBDT | whole trip |
| | [41] | | 23062 trips | GPS | Walk, Bike, Bus, Car, Taxi, Train, Subway | DNN | whole trip |
| | [42] | Private | 4685 trips | GPS | Walk, Bike, Ebike, Car, Bus | BN | whole trip |
| | [43] | | 6.2 h | GPS, GIS | Still, Walk, Bike, Bus, Car, Train | NB, BN, DT, RF, ML | 30 s |
| | [44] | | 30000 trips | GPS, GIS | Walk, Bike, Bus, Car, Train | SVM | whole trip |
| | [45] | | 900 h | GPS | Walk, Bike, Bus, Car, Train, Subway | SVM | 180 s |
| | [46] | | 340 trips | GPS, GIS | Walk, Bus, Car, Rail, Subway | Hierarchical decision | whole trip |
| | [47] | | 114 trips | GPS | Walk, Bus, Car | NN | whole trip |
| $\mathcal{H}$ | [48] | Private | NA | GPS, Acc | Walk, Run, Bike, Bus, Motorcycle, Car, Tain, Tram, Metro, Light rail | BBN | 60 s |
| | [49] | | 120 h | GPS, Acc | Walk, Run, Bike, Vehicle | CHMM, DT+DHMM | 1 s |
| | [50] | | 355 h | GPS, Acc | Walk, Bike, Motorcycle, Bus, Car, Train, Tram, Subway | Ensemble+HMM | 10 s |
| | [51] | | NA | GPS, Acc, Mic | Still, Walk, Run, Bike, Vehicle | DT, MG, SVM, NB, GMM, MDP | 2-60 s |
| | [52] | | 266 h | GPS, Acc | Still, Walk, Bike, Motorcycle, Bus, Car, Train, Tram, Subway | RSM + HMM | 5 s |
| | [53] | | 22 h | GPS, Acc | Still, Walk, Bike, Vehicle | SVM | 5 s |

BBN - Bayesian belief network; BN - Bayesian network; CHMM - coupled hidden Markov model; CRF - conditional random field; DHMM - discrete hidden Markov model; DNN - deep neural network; DT - decision tree; GBDT - gradient boosted decision tree; GMM - Gaussian mixture model; HMM - hidden Markov model; KNN - k-nearest neighbour; NB - naive Bayesian; MDP - Markov decision process; ML - multilayer perception; NN - neural network; QDA - quadratic discriminant analysis; RF - random forest; RSM - random subspace method; SVM - support vector machine.

varying from 1 second to 18 seconds, aiming at real-time decision. The most widely used choice is around 5 seconds. An exception was reported in [37], which used a barometer sensor alone to predict the mode of transportation within a window size of 200 seconds. Location based approaches usually employ a long window varying from several minutes to tens of minutes or even the entire trip. In the latter case, the decisions are made offline with applications in travel surveys. Hybrid approaches target real-time decision by combining inertial and GPS sensors, and thus prefer a short window with sizes similar to the ones used in inertial based approaches.

*Classifier*. Various classifiers have been employed for the recognition task. Decision tree (DT), K-nearest neighbour (KNN), support vector machine (SVM) and naive Bayesian (NB) are the most frequently used classifiers. Several schemes were proposed to improve the classification performance, such as ensemble classifiers, multi-layer classifiers, and post-processing. AdaBoost [22], [40] and random forest (RF) [24], [29], [32], [35], [40], [43], [50] ensemble a set of simple classifiers for the optimal decision. Multi-layer classifiers typically perform a coarse-grained distinction between pedestrian and motorized transportation in the first tier, and then perform a fine-grained classification in the subsequent tiers [22], [25]–[27], [53]. Post-processing can reduce the classification error effectively by using a voting scheme which exploits the temporal correlation between consecutive frames [22], [28] or using a hidden Markov model (HMM) to capture the transition probability between different classes [48]–[50], [52]. Long-term features were computed using the information from whole trip to improve the classification accuracy in short segments [25]. Deep learning, which attracts significant interests in the machine learning community, was recently applied to the transportation mode recognition task [21], [41]. For

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



**TABLE 2.** Data channels derived from the smartphone sensors.

| Sensor | Data channel | Reference |
|---|---|---|
| Inertial | Magnitude of accelerometer data | [20], [22]–[24], [27], [28], [30]–[36], [49]–[54] |
| | Horizontal and vertical magnitude of accelerometer data | [23], [25], [33] |
| | Calibrated three axes of accelerometer data | [26], [29], [48] |
| | Magnitude of gyroscope data | [20], [22], [32], [35] |
| | Magnitude of magnetometer data | [20], [22], [26], [35] |
| | Magnitude of barometer data | [26], [37] |
| GPS | Speed | [38]–[40], [42]–[53] |
| | Acceleration | [38]–[40], [42]–[47], [50], [52] |
| | Turn angle | [40], [43] |
| | Trajectory | [41] |

**TABLE 3.** Time domain ($\mathcal{T}$) and frequency domain ($\mathcal{F}$) features computed on the data channel derived from inertial sensors.

| Type | Feature | Reference |
|---|---|---|
| $\mathcal{T}$ | Mean | [20], [22]–[26], [28]–[33], [35], [37], [48]–[51], [54] |
| | Standard deviation (variance) | [20], [22]–[26], [28], [31]–[34], [36], [48]–[51], [54] |
| | Mean crossing rate | [20], [23]–[25], [28], [33], [34], [51] |
| | Energy | [24], [25], [31], [34] |
| | Auto correlation, Kurtosis, Skewness | [25] |
| | Min, Max | [25], [26], [29], [32], [34] |
| | Median | [25], [35] |
| | Range | [24], [25] |
| | Third quartile | [23], [27], [28], [33] |
| | Quantiles 5, 25, 50, 75, 90 squared sum above/below these quantile | [27] |
| | Interquartile range | [24], [25], [33], [35] |
| $\mathcal{F}$ | Frequency with highest FFT value | [20], [22], [23], [25], [26], [28], [33] |
| | Ratio between the first and second highest FFT peaks | [20], [22] |
| | Mean, Standard deviation | [54] |
| | DC of FFT | [25], [26] |
| | All the FFT values | [31], [36], [50], [52] |
| | Sum and std in the frequency 0-2 Hz | [23], [28] |
| | Ratio between the energy in frequency 0-2 Hz and the whole band | [23], [28] |
| | Sum and std in the frequency 2-4 Hz | [23], [28] |
| | Ratio between the energy in frequency 2-4 Hz and the whole band | [23], [28] |
| | Energy at 1, 2, $\cdots$, 10 Hz | [25], [49] |
| | Energy at [0, 1], [1, 3], [3, 5], [5, 16] Hz, and the ratio between them | [51] |

performance evaluation, two objective measures are widely used: the F1-score and the recognition accuracy computed from the confusion matrix.

### *B. FEATURES FOR TRANSPORTATION MODE RECOGNITION*

Feature computation is the key for transportation mode recognition. Most publications report a different scheme to compute features from the multimodal sensor data. To help understand the state of the art, we first summarize the data channels that are used to compute features from various modalities (Table 2), and then summarize specific features that are computed in each data channel (Table 3 and 4).

Table 2 lists the data channels that are used to compute features from inertial and GPS sensors. Accelerometer, which measures the acceleration along three device axes, is the most favoured modality among inertial sensors. Since the pose and orientation of the mobile device is typically unknown, several approaches have been proposed to extract orientation independent information, e.g. by computing the magnitude which combines acceleration from three axes [20], [22], [23], [27], [28], [30]–[36], [49]–[54], by decomposing the magnitude along a vertical and horizontal earth coordinate system [23], [25], [33], or by projecting the raw acceleration of the three device axes into a 3D earth coordinate system [26], [29], [48]. The magnitudes of the data from other modalities, including gyroscope [20], [22], [32], [35], magnetometer [20], [22], [32], [35] and barometer [26], [37], have also been used for feature computation.

**TABLE 4.** Features computed on the data channel derived from the GPS sensors.

| Feature | Reference |
|---|---|
| Mean | [38]–[40], [42]–[53] |
| Stand deviation (variance) | [38]–[40] |
| Sinuosity | [40] |
| Range; Interquartile range | [40] |
| Max | [47] |
| Quantile 25 and 75 | [40], [46] |
| Quantile 95 | [42], [46] |
| Three maximum values | [38]–[40] |
| Three minimum values | [40] |
| Autocorrelation; Kurtosis; Skewness | [40] |
| Heading change rate | [40], [43] |
| Velocity change rate; Stop rate | [40] |

Table 3 lists the specific features that can be computed in each inertial sensor data channel (Table 2), which can be time-domain and frequency-domain. The time-domain features are computed based on a frame of samples while the frequency-domain features are computed based on the fast Fourier transform (FFT) of a frame of samples. Mean, standard deviation, mean crossing rate, and energy are among the most popular time-domain features. The quantile value and quantile range of the samples in a frame are widely used to represent the minimum, maximum, median value and interquartile range of the samples in a frame. However, the choices on which quantile appear to be rather ad-hoc among the literature. Statistical measures such as auto-correlation, kurtosis, and skewness are less frequently reported. The most used frequency domain feature is the frequency with the highest energy peak. The energy in different frequency bands is a widely used feature. However, the choices of a specific subband appears to be rather ad-hoc among the literature. For instance, the reference [25], [49] considered the energy specifically at 1 Hz, 2 Hz, $\cdots$, 10 Hz, while the reference [51] considered the energy between 0 and 1 Hz, 1 and 3 Hz, 3 and 5 Hz, 5 and 16 Hz. Some statistical features such as the ratio between the first and the second FFT peaks, the mean and standard deviation of the FFT coefficients have also been suggested.

Table 2 also lists the data channels that can be derived from the GPS sensors, including speed, acceleration, turning

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



angle and trajectory. These data channels are inferred from the change of GPS location over time. Table 4 lists the specific features that can be computed in these data channels. GPS features are usually computed in the time domain only. Mean and standard deviation are two most popular features computed from speed, acceleration and turn angle. Different choices of quantile and quantile ranges (e.g. max, quartile, and interquartile range) and statistics (e.g. kurtosis and skewness) are widely used features computed from speed, acceleration and turn angle. Several advanced features including heading change rate, stop rate, and velocity change rate are also proposed and computed for a single trip. For hybrid approaches, which compute GPS features in a short window, only mean and standard deviation of speed or acceleration are used [48]–[53].

To summarize, while transportation mode recognition has been investigated intensively and with great advances reported in recent years, the work of various research groups was conducted in a rather isolated way and does not show close inter-connection in the research community. Each work appears to define its own transportation mode classification problem (e.g. the number of activities considered), and proposes a solution with different parameters (e.g. window size, sensor modality, classifier), and often verified with ad-hoc datasets which are not public available. A fair comparison of results between different groups is very difficult. As the number of publications increases, this obviously holds back research advances in this area as it prevents systematic comparative evaluation of novel methods or sensors.

The research community has proposed a large number of features for transportation mode recognition. While effective, these features appear to be defined in as rather ad-hoc manner and they are computed from different modalities. In particular, there is few unity in the literature on the time-domain quantiles and sub-band energy to employ as features.

## III. SHL DATASET

The University of Sussex-Huawei Locomotion (SHL) dataset is a major outcome of our large-scale longitudinal data collection campaign, which collected 2812 hours of labeled data over a period of 7 months which corresponds to 17,562 km in the south-east of the UK including London [18], [19]. The SHL dataset was recorded by three participants engaging in eight transportation and locomotion activities in real-life settings: Still, Walk, Run, Bike, Car, Bus, Train and Subway. Each participant carried four Huawei Mate 9 smartphones at four body positions simultaneously: in the hand, at the torso (located in a shirt or jacket pocket or a torso strap), at the hip, in a backpack or handbag (Fig. 2). Each smartphone logged the data of the 16 sensors available in the smartphone, including inertial sensors, GPS, ambient pressure sensor, ambient humidity, etc. The data from four smartphones leads to a total duration of $4 \times 703 = 2812$ hours. In addition to the smartphones, each participant wore a front-facing camera to record images of the environment during the journey, which


**FIGURE 2.** A participant wearing the four smartphones and a camera during data collection.

**TABLE 5.** Characteristics of the SHL dataset.

| | |
|---|---|
| User | 3 |
| Body position | Hand, Torso, Hip, Bag |
| Modalities (sampling rate) considered in this paper | GPS (1 Hz), Accelerometer (100 Hz), Gyroscope (100 Hz), Magnetometer (100 Hz) |
| Modalities not considered in this paper | Linear accelerometer, Orientation, Gravity, Barometer, Satellite, Ambient light, Battery, Temperature, Wifi, GSM, Ambient sound, Image, Google API |
| Transportation activity | Still, Walk, Run, Bike, Car, Bus, Train, Subway |
| Total duration | $4 \times 703 = 2812$ hours |

was used to precisely annotate the activities of the user. Table 5 indicates the characteristics of the dataset.

Fig. 3 depicts (a) the duration of each transportation activity performed by the three participants and (b) the duration of the transportation activities where the GPS data is available. The GPS information might not always be available during the journey, e.g. when the user is taking a subway or is staying inside a building. In the dataset, we regard a segment as 'GPS off' if this segment has no GPS information available for more than 10 seconds. We refer to the case (a) Dataset-E, i.e. the entire dataset is used, and the case (b) as Dataset-IG, i.e. the subset of Dataset-E where data from the GPS sensor is available. The total amount of data are 2812 and 2036 hours, respectively.

The SHL dataset is well suited to enable systematic comparative evaluations of recognition methods. It contains

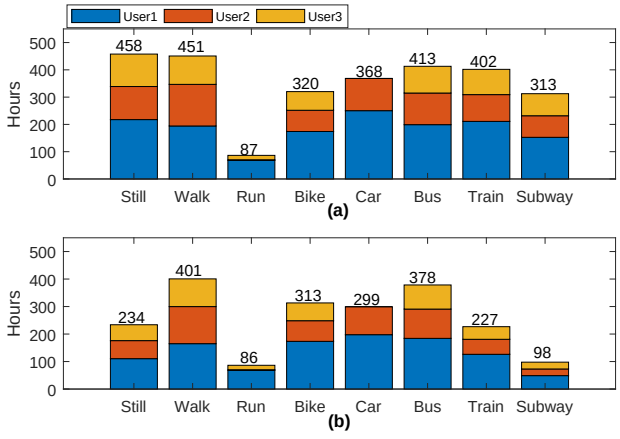


**FIGURE 3.** (a) Amount of data (Dataset-E) collected for each of the eight transportation activities by the three users. (b) Amount of data (Dataset-IG) where GPS is available.

all the modalities ever used in the 34 related work and contains all of the activity classes in 25 out of 34 related work. The duration of the dataset is much longer than any dataset reported in the literature with both inertial and GPS data. The dataset contains data recorded at multiple body positions and by multiple users. Therefore, this dataset allows to replicate the majority (25 out of 34) of the experiments reviewed in the related work.

For clarity, we introduce the following naming schemes for the transportation and locomotion modes: S1-Still, W2-Walk, R3-Run, B4-Bike, C5-Car, B6-Bus, T7-Train, S8-Subway. W2, R3 and B4 belong to the pedestrian activity of the user, where W2 and R3 can be categorized as foot activities. C5, B6, T7 and S8 belong to a family of vehicular transportation, where C5 and B6 can be categorized as road transportation and T7 and S8 categorized as rail transportation.

## IV. RECOMMENDED TRANSPORTATION MODE RECOGNITION TASKS

In order to enable reproducible research in transportation mode recognition it is important that the recognition scenarios are well defined. However, it is also important that they suit existing and foreseeable demands from different applications. In this section we propose a list of generalized transportation mode recognition tasks that aim to cover most application scenarios considered in the literature. As shown in Table 6 these tasks consists of 12 subgroups (scenarios) based on the eight classes in the SHL dataset: S1-Still, W2-Walk, R3-Run, B4-Bike, C5-Car, B6-Bus, T7-Train, S8-Subway.

This subgrouping scheme merges one or more activities together into a new class based on application interests. For instance, Pedestrian (Walk, Run, Bike), Vehicle (Bus, Car, Train, Subway), Foot (Walk, Run), Road vehicle (Bus, Car), Rail vehicle (Train, Subway), are new classes merging existing activities. A detailed description of the 12 scenarios is given below.

- Scenario 1 is based on the physical activity of the user and categorizes the eight activities into Physically Active (Walk, Run and Bike) and Inactive (Still and Vehicle).
- Scenario 2 is based on the power source (human-powered or machine-powered) and categorizes the eight activities into Still, Pedestrian (Walk, Run and Bike) and Vehicle (Car, Bus, Train and Subway).
- Scenarios 3 and 4 merge the four vehicle activities into a new group Vehicle. Scenario 3 additionally merges Walk and Run into Foot.
- Scenarios 5 and 6 categorize the four vehicle activities into Road vehicle (Car and Bus) and Rail vehicle (Train and Subway). Scenario 5 additionally merges Walk and Run into Foot.
- Scenarios 7 and 8 categorize the four vehicle activities into Private vehicle (Car) and Public vehicle (Bus, Train and Subway). Scenario 7 additionally merges Walk and Run into Foot.
- Scenarios 9 and 10 categorize the four vehicle classes into Private road vehicle (Car), Public road vehicle (Bus), and Rail vehicle (Subway and Train). Scenario 9 additionally merges Walk and Run into Foot.
- Scenario 11 only merges Walk and Run into Foot. Scenario 12 does not have any subgrouping, i.e. with the original eight classes contained in the SHL dataset.

Table 6 links the 12 scenarios to related literature in the first column. These 12 scenarios cover most transportation mode recognition tasks considered in the literature (25 out of 34 related work) and link closely to the remaining ones which contain more activities than the SHL dataset, e.g. Motorcycle [20]–[22], [36], [48], [50], [52], E-bike [42], Boat and Plane [32]. Some of these scenarios can be used to encourage a more ecologically friendly or physically active lifestyle, or provide appropriate contextual information.

When developing a system to automatically recognize transportation modes it is important to evaluate it according to its final usage patterns. We thus propose to evaluate the recognition performance of the 12 scenarios from three perspective: user-independent, position-dependent, and time-invariant evaluation (Table 7).

Generally, a recognition system should work regardless of whom is using it. However, human motion dynamics varies between users due to physical characteristics and habits. For instance, different users may have different gait styles and ideal walking or jogging speed, or may engage in different activities when they are in public transport (e.g. reading a book, tapping to music, etc.). User-independent activity recognition aims to design recognition systems that will generalize well to new users [57]. We divide the dataset based on the three users and evaluate the performance with a leave-one-user-out crossvalidation, e.g. training with the data from User 2 and User 3 and testing with the data from User 1.

A recognition system based on smartphones should ideally operate regardless of where the users carry their phone, i.e. it should be position-independent. We divide the dataset based on the four positions and evaluate the performance with a leave-one-position-out cross-validation, e.g. training with the data from Torso, Hip and Bag and testing with the data from Hand.

Finally, a system should keep operate over time, despite possible changes in behaviour (e.g. due to injury, different preferences or habits), i.e. it should be time-invariant. With data collected over the course of 7 months, we can assess this in the SHL dataset through a leave-one-period-out cross-validation, where a period is composed of the data of consecutive days of recordings. Specifically, we divide the dataset into four periods based on the recording dates of the three users, and perform training with three periods and testing with the remaining period. Table 8 presents the number of recording days in each period, and the duration of each transportation activity within each period.

In related work various modalities were employed for transportation mode recognition, where accelerometer and GPS are the most used ones. Historically, the earlier phones

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



**TABLE 6.** Subgrouping based on the eight classes in the SHL dataset: S1-Still, W2-Walk, R3-Run, B4-Bike, C5-Car, B6-Bus, T7-Train, S8-Subway.

| Reference | Subgroups | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| [22], [25]–[27], [53] | **Scenario 1** | (W2-B4)<br>Active | (S1, C5-S8)<br>Inactive | | | | | |
| [37] | **Scenario 2** | S1<br>Still | (W2-B4)<br>Pedestrian | (C5-S8)<br>Vehicle | | | | |
| [53] | **Scenario 3** | S1<br>Still | (W2, R3)<br>Foot | B4<br>Bike | (C5-S8)<br>Vehicle | | | |
| [23], [27], [33], [35], [49], [51] | **Scenario 4** | S1<br>Still | W2<br>Walk | R3<br>Run | B4<br>Bike | (C5-S8)<br>Vehicle | | |
| [29], [30], [34] | **Scenario 5** | S1<br>Still | (W2, R3)<br>Foot | B4<br>Bike | (C5, B6)<br>Road | (T7, S8)<br>Rail | | |
| / | **Scenario 6** | S1<br>Still | W2<br>Walk | R3<br>Run | B4<br>Bike | (C5, B6)<br>Road vehicle | (T7, S8)<br>Rail vehicle | |
| / | **Scenario 7** | S1<br>Still | (W2, R3)<br>Foot | B4<br>Bike | C5<br>Private road vehicle | (B6-S8)<br>Public vehicle | | |
| [24] | **Scenario 8** | S1<br>Still | W2<br>Walk | R3<br>Run | B4<br>Bike | C5<br>Private road vehicle | (B6-S8)<br>Public vehicle | |
| [23], [28], [38], [39], [43], [44], [47], [42]* | **Scenario 9** | S1<br>Still | (W2, R3)<br>Foot | B4<br>Bike | C5<br>Private road vehicle | B6<br>Public road vehicle | (T7, S8)<br>Rail vehicle | |
| [26] | **Scenario 10** | S1<br>Still | W2<br>Walk | R3<br>Run | B4<br>Bike | C5<br>Private road vehicle | B6<br>Public road vehicle | (T7, S8)<br>Rail vehicle |
| [25], [40], [41], [45], [46], [32]*, [36]*, [50]*, [52]* | **Scenario 11** | S1<br>Still | (W2, R3)<br>Foot | B4<br>Bike | C5<br>Car | B6<br>Bus | T7<br>Train | S8<br>Subway |
| [20]*, [21]*, [22]*, [48]* | **Scenario 12** | S1<br>Still | W2<br>Walk | R3<br>Run | B4<br>Bike | C5<br>Car | B6<br>Bus | T7<br>Train | S8<br>Subway |

The superscript * denotes that the referenced work contains more transpiration activities than the SHL dataset.

only comprised an accelerometer as a motion sensor and thus a large amount of work focused on transportation mode recognition using this sensor only. As time evolves, multimodal motion sensors (accelerometer, gyroscope and magnetometer) were integrated into a single smartphone chip. In recent years an increasing number of work performs transportation mode recognition using multimodal sensors. Because not all the work use the same sensor configuration, we need to evaluate the recognition performance using combination of sensors which form a superset of the related work. To this end, we propose the following six group of modalities as a combination of accelerometer, gyroscope, magnetometer and GPS: A (Acc), AG (Acc + Gyr), AGM (Acc + Gyr + Mag), P (GPS), AP (Acc + GPS), AGMP (Acc + Gyr + Mag + GPS). First, acceleration and GPS are the most common sensors in smartphones and we are interested to investigate the recognition performance with these two modalities alone (A and P) and the combination of them (AP). Second, the energy usage of a gyroscope is significantly higher (order of magnitude) than that of an accelerometer, which thus essentially comes for free if the gyroscope is turned on. The magnetometer uses about twice the energy than the gyroscope. When the magnetometer is enabled, the gyroscope and accelerometer can be enabled with little extra energy usage. We thus propose to use the combinations AG and AGM. GPS uses an order of magnitude more than the magnetometer. If we turn on GPS, the other motion sensors can be enabled as well without significant energetic impact. We thus propose to evaluate the combination AGMP.

Table 7 lists 792 recognition tasks, as a combination of a recognition scenario, out of the 12 suggested, a leave-one-out scheme to assess user, position or temporal independence, and a group of sensor modalities.

GPS is not always available in the entire dataset (see Fig. 3 and Table 8). When evaluating modalities A, AG and AGM, we use the entire dataset, i.e. the Dataset-E. When evaluating the modalities P, AP and AGMP, we use the dataset where the GPS is available, i.e. Dataset-IG (see Fig. 3 and Table 8). For ease of comparison between the six groups of modalities, we also use Datase-IG to evaluate A, AG and AGM.

For performance evaluation, we opt for two measures, recognition accuracy and F1 score, which are widely used in the literature. While recognition accuracy gives an intuitive indication of the performance, F1 score can better handle imbalance datasets between classes. The two measures can be computed from the confusion matrix between the output labels and the ground-truth labels. Let $M_{ij}$ be the $(i, j)$-the element of the confusion matrix. It represents the number of samples originally belonging to class $i$ which are recognized as class $j$. Let $C$ be the number of classes. The accuracy (R) and the F1-score (F) are defined as follows.

$$R = \frac{\sum_{i=1}^{C} M_{ii}}{\sum_{i=1}^{C}\sum_{j=1}^{C} M_{ij}}, \tag{1}$$

$$recall_i = \frac{M_{ii}}{\sum_{j=1}^{C} M_{ij}}, \quad precision_j = \frac{M_{jj}}{\sum_{i=1}^{C} M_{ij}}, \tag{2}$$

$$F = \frac{1}{C}\sum_{i=1}^{C} \frac{2 \cdot recall_i \cdot precision_i}{recall_i + precision_i}. \tag{3}$$

## V. FEATURE ANALYSIS

The large amount of data in the SHL dataset allows us to conduct a thorough analysis to investigate the ability of a large set of features to distinguish between any two transportation activities. To this end, we first define a

**TABLE 7.** Recommended transportation mode recognition tasks using SHL dataset.

| Leave-one-X-out cross-validation | | Scenario | Modality |
|---|---|---|---|
| User-independent<br>X = user | User 1 | 1-12 | A (Acc)<br>AG (Acc + Gyr)<br>AGM (Acc + Gyr + Mag)<br>P (GPS)<br>AP (Acc + GPS)<br>AGMP (Acc + Gyr + Mag + GPS) |
| | User 2 | | |
| | User 3 | | |
| Position-independent<br>X = position | Hand | | |
| | Torso | | |
| | Hip | | |
| | Bag | | |
| Time-invariant<br>X = period | Period 1 | | |
| | Period 2 | | |
| | Period 3 | | |
| | Period 4 | | |

**TABLE 8.** Division of the SHL dataset based on the recording days.

| **Recording days** | Period 1 | Period 2 | Period 3 | Period 4 |
|---|---|---|---|---|
| User 1 (82 days) | 1-15 | 16-44 | 45-62 | 63-82 |
| User 2 (40 days) | 1-12 | 13-16 | 17-32 | 33-40 |
| User 3 (30 days) | 1-7 | 8-12 | 13-22 | 23-30 |
| **Activity duration / GPS available (hours)** | | | | |
| Still | 126 / 68 | 104 / 49 | 103 / 61 | 124 / 55 |
| Walk | 110 / 99 | 110 / 96 | 115 / 104 | 115 / 102 |
| Run | 22 / 22 | 22 / 22 | 21 / 21 | 21 / 21 |
| Bike | 79 / 75 | 79 / 78 | 93 / 92 | 70 / 69 |
| Car | 120 / 98 | 121 / 94 | 90 / 76 | 37 / 31 |
| Bus | 115 / 106 | 103 / 93 | 99 / 92 | 96 / 88 |
| Train | 76 / 47 | 85 / 43 | 100 / 57 | 141 / 80 |
| Subway | 59 / 16 | 65 / 20 | 65 / 19 | 124 / 43 |

set of features that can be computed from the various modalities, and then perform a discriminablity analysis based on the mutual information between these features and the transportation modes. Finally we employ a filter-based feature selection algorithm employing a maximum-relevance-minimum-redundency (MRMR) criteria [58] to preselect a subset of features, which are subsequently used to establish the baseline performance measures for the tasks identified in the previous section.

### A. FEATURE EXTRACTION

We compute the features within a short-time window of 5.12 seconds, which is the most common duration we identified in Table 1. As shown in the state-of-the-art analysis in Sec. II-B and Table 4, most GPS features are computed in long temporal intervals except the mean speed and mean acceleration. As we are interested in just-in-time context recognition and thus work with short frames, we only compute these two features for the GPS data. For this reason, the analysis of GPS features will not be considered in this section and will be limited to the data coming from the three inertial sensors: accelerometer, gyroscope and magnetometer. For each modality we use the magnitude of the data channel for feature computation. The magnitude has been widely used in the literature and is robust to the variation of device orientation (Table 2).

Through related work analysis, we noticed that while a variety of features have been proposed for transportation mode recognition, the choices of these features appear to

**TABLE 9.** Feature analysis: subband ($\mathcal{E}$) and quantile ($\mathcal{Q}$) features, and the remaining time-frequency domain ($\mathcal{T}$+$\mathcal{F}$) features.

| **Type** | **Features** | **Dimension** |
|---|---|---|
| $\mathcal{E}$ | Energy and energy ratio with scan width 1 Hz and skip 0.5 Hz | 198 |
| | Energy and energy ratio with scan width 2 Hz and skip 1 Hz | 98 |
| | Energy and energy ratio with scan width 3 Hz and skip 1 Hz | 96 |
| | Energy and energy ratio with scan width 4 Hz and skip 1 Hz | 94 |
| | Energy and energy ratio with scan width 5 Hz and skip 1 Hz | 92 |
| | Energy and energy ratio with scan width 10 Hz and skip 1 Hz | 82 |
| | Energy and energy ratio with scan width 15 Hz and skip 1 Hz | 72 |
| | Energy and energy ratio with scan width 20 Hz and skip 1 Hz | 62 |
| | Energy and energy ratio with scan width 25 Hz and skip 1 Hz | 52 |
| | Total | **846** |
| $\mathcal{Q}$ | Quartiles: [0, 5, 10, 25, 50, 75, 90, 95, 100] | 9 |
| | Pairwise quartile range for the 9 quartiles | 36 |
| | Total | **45** |
| $\mathcal{T}$ | Mean, standard deviation, energy | 3 |
| | Mean crossing rate | 1 |
| | Kurtosis and Skewness | 2 |
| | Highest auto correlation value and offset | 2 |
| $\mathcal{F}$ | DC component of FFT | 1 |
| | Highest FFT value and frequency | 2 |
| | Ratio between the highest and the second FFT peaks | 1 |
| | Mean, standard deviation | 2 |
| | Kurtosis and skewness | 2 |
| | Energy | 1 |
| | Total | **17** |

be rather ad-hoc, especially on the subband energy and the quantile range. It would be interesting to find out which feature provides the most distinctive power for the recognition task. To perform an exhaustive evaluation, we compute all the features that are listed in the literature (Table 3) and we additionally compute a set of quantile and subband features. Table 9 lists the features to be computed, which can be categorized into three families: subband energy ($\mathcal{E}$), time-domain quantile ($\mathcal{Q}$), and the remaining time-domain and frequency-domain ($\mathcal{T}$+$\mathcal{F}$) features.

A subband is usually defined with two parameters: centre frequency $\omega_c$ and bandwidth $\omega_b$. The frequencies in a subband is thus given by $\omega \in [\omega_c - \frac{\omega_b}{2}, \omega_c + \frac{\omega_b}{2}]$. Instead of evaluating the ad-hoc subband features defined in the literature, we propose to systematically compute a set of subband features with all possible parameters of $\omega_c$ and $\omega_b$. The highest frequency of the data is 50 Hz as the sampling rate is 100 H. We consider the following bandwidth: $\omega_b \in \{1, 2, 3, 4, 5, 10, 15, 20, 25\}$ Hz. For each bandwidth $\omega_b$, we vary the centre frequency from $\frac{\omega_b}{2}$ to $50 - \frac{\omega_b}{2}$ with a step of 1 Hz. For the bandwidth $\omega_b = 1$ Hz the center frequency is increased with a step of 0.5 Hz. For each subband, we consider two types of features: the absolute energy and the energy ratio. Let $\{S_1, \cdots, S_K\}$ represent the $K = 257$ FFT coefficients of a frame of data and let $k_L$ and $k_H$ denotes the indices of the lower and upper frequencies of a subband $[\omega_c - \frac{\omega_b}{2}, \omega_c + \frac{\omega_b}{2}]$, the two features are defined as

$$f_{\text{subegr}} = \sum_{k=k_L}^{k_H} |S_k|^2, \tag{4}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



$$f_{\text{subratio}} = \frac{\sum_{k=k_L}^{k_H} |S_k|^2}{\sum_{k=1}^{K} |S_k|^2}. \quad (5)$$

Finally we obtain 846 features in the set $\mathcal{E}$ as shown in Table 9.

A quantile range $[q_L, q_H]$ is defined as $s(q_H) - s(q_L)$, the difference between two percentile values $s(q_L)$ and $s(q_H)$ of a frame of samples $s$. Instead of evaluating the ad-hoc quantile and quantile-range features defined in the literature, we propose to systematically compute a set of quantile features with a list of possible parameters of $q_L$ and $q_H$. We consider the following 9 quantile values $q_L, q_H \in \{0, 5, 10, 25, 50, 75, 90, 95, 100\}$ with $q_L \leq q_H$. This results in 9 quantiles with $q_L = q_H$ and 36 quantile ranges with $q_L < q_H$. Finally we obtain 45 features in the set $\mathcal{Q}$ as shown in Table 9.

We include all the time-domain ($\mathcal{T}$) and frequency-domain ($\mathcal{F}$) features, excluding the quantile and subband features, that are listed in Table 3, which yields 17 features containing 8 elements in the time domain and 9 in the frequency domain.

With the proposed scheme, we compute $17 + 45 + 855 = 908$ features for each modality and thus $3 \times 908 = 2724$ features per frame of inertial sensor data in total. The frames are obtained by sliding a window of 5.12 seconds with 2.56 s overlap on the entire dataset. This yields 3.95 million frames, each containing 2724 features.

### B. FEATURE ANALYSIS BASED ON MUTUAL INFORMATION

Given so many features computed in each data frame, we are interested in finding the answers to three questions: which *modality*, which *quantile* range, and which *subband* is most informative to distinguish between transportation modes.

Mutual information (MI) is widely used to measure the relevance between features and target classes, and also the dependency between features [58]–[60]. Given two variables $x$ and $y$, the probability density functions (pdf) $p(x)$ and $p(y)$, and the joint pdf $p(x, y)$, the mutual informational is defined as

$$I(x; y) = \int \int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dxdy. \quad (6)$$

The MI $I(x; y)$ lies in the range [0, 1], with a value close to 1 indicating a strong dependency between two variables and a value of 0 indicating independence between them. For a specific recognition problem with a feature $f$ and a set of classes $C$, a higher MI value $I(f, C)$ indicates a stronger ability of the feature to distinguish between these classes [59].

We employ mutual information as a measure to investigate the discriminablity of each feature on any two transportation activities. Given the eight activity classes in the SHL dataset there are 28 pair-wise combinations of any two. We compute the mutual information between each feature and each class pair. When computing mutual information, the pdf of the feature variable in Eq. (6) is approximated with the histogram over all (3.95 million) instances. Specifically, $p(x)$ or $p(y)$ is approximated with a 1-dimension histogram with a fixed number of 500 bins; $p(x, y)$ is approximated with a 2-dimension histogram with 200 bins at each dimension.

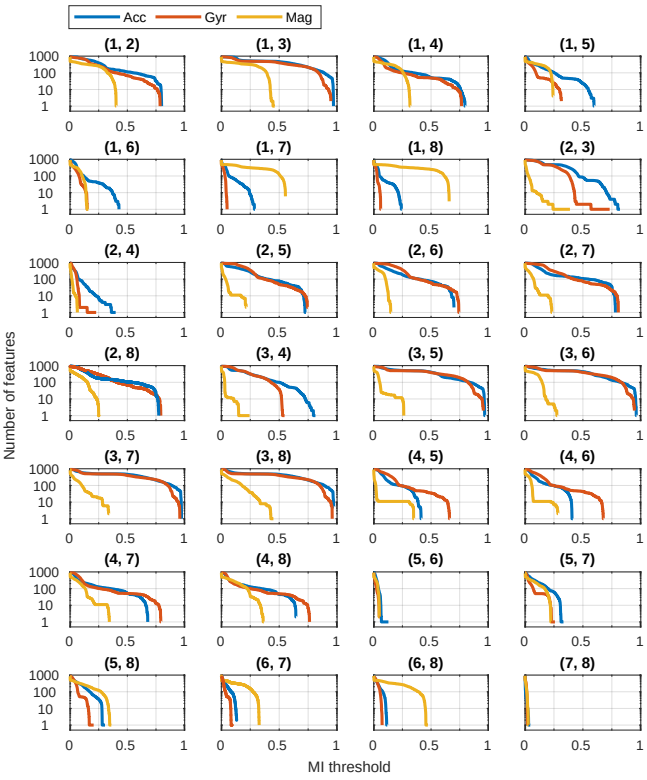


**FIGURE 4.** For each modality (accelerometer, gyroscope and magnetometer) we extract 908 features and compute the MI between each feature and the 28 pair-wise combinations of the eight classes: S1-Still, W2-Walk, R3-Run, B4-Bike, C5-Car, B6-Bus, T7-Train, S8-Subway. The figure shows the number of features from each modality that presents an MI value above a specified threshold.

For convenience, we use the notation (S1/W2 vs R3/B4) to represent the task of distinguish between two classes (S1, R3), (S1, B4), (W2, R3), or (W2, B4).

#### 1) Modality

For a specific recognition problem, a feature with a higher MI value usually indicates a stronger ability to separate the target classes. We thus use the number of features with a high value of MI (above a threshold $I_T$) contained in one single modality (accelerometer, gyroscope, or magnetometer) to indicate the significance of this modality to the recognition task. If we do not consider the redundancy of the features in the same modality, the more high-MI features the more important this modality is. For each modality (with 908 features) and each pair of classes (the 28 pair-wise combination from eight classes), we compute the number of features with MI above a threshold $I_T$. We plot in Fig. 4 how this number varies in function of the threshold $I_T$. The following observations can be made regarding the significance of each modality.

All the three modalities present very few features with high MI values for two pairs (C5 vs B6) and (T7 vs S8).

The likely explanation for this is that the motion pattern of Car and Bus are very similar specially in short time frames, and so do the Train and Subway. For (T7 vs S8) no feature from the three modalities present an MI value higher than 0.05, which implies the two classes are almost indistinguishable with a single feature. For (C5 vs B6), all the features from gyroscope and magnetometer shows an MI value below 0.05, while accelerometer has less than 10 features with MI between 0.05 and 0.1. This implies that accelerometer provides more distinctive power than the other two modalities for separating C5 and B6, although making this distinction appears to be comparatively more difficult.

For each of the remaining 26 pairs, either one or several of the three modalities can provide features with a high MI value. Accelerometer and gyroscope show similar significance curves across many class pairs, such as (S1 vs W2/R3/B4) and (W2/R3 vs C5/B6/T7/S8). These two modalities provide a similar number of features with high MI values (e.g. $> 0.7$) when distinguishing between Still (S1) and pedestrian (W2/ R3/B4), and between foot (W2/R3) and vehicles (C5/B6/ T7/S8). Accelerometer provides more high-MI features than gyroscope for most of the remaining pairs, e.g. when distinguishing between Still (C1) and four vehicles (C5/B6/ T7/S8), and also between the three pedestrian activities (W2 vs R3 vs B4). Gyroscope provides more high-MI features than accelerometer when distinguishing Bike (B4) and the four vehicles. This is possibly because the Bike activity introduces more rotational motions than vehicles. Both accelerometer and gyroscope provides very few high-MI features when distinguishing between the four vehicles (i.e. C5 vs B6 vs T7 vs S8).

Magnetometer usually provides much less high-MI features than accelerometer and gyroscope for most class pairs, because the ambient magnetic field is not closely related to the human activity in open-spaces, where there is little magnetic disturbance due to the presence of surrounding metals. However, the magnetometer provides significantly more high-MI features than the other two modalities when distinguishing between Still (S1) and rail transportation (T7/S8), and between driving (C5/B6) and rail (T7/S8). This is an interesting observation that has not been reported in the previous literature. One possible explanation could be the influence of metal casing of the train and subway.

#### 2) Subband energy

Fig. 5 visualizes the MI values between subband features (family $\mathcal{E}$) from the three modalities and the 28 class pairs. Each subfigure contains 28 panels corresponding to the 28 class pairs. Each panel consists of two parts: the upper block shows the MI between the energy-ratio features and the class pairs; the lower block shows the MI between absolute-energy features and the class pairs. The x-axis denotes the center frequency of the subband which varies from 0 to 50 Hz, while the y-axis denotes the bandwidth, which varies from 1 to 25 Hz. Based on the MI values we can easily find out which subband provides a higher discriminablity between the target classes.

For accelerometer and gyroscope in Fig. 5(a) and (b), the lower block (absolute energy) provide more high-MI features than the upper block (energy ratio). For accelerometer, most high MI values are observed in low frequency, especially between 0 and 10 Hz. For gyroscope, most high MI values are observed in low frequency, especially between 5 and 10 Hz and some class pairs, e.g. (B4 vs C5/B6/T7/S8), present high MI values in the frequency band between 0 and 5 Hz. For accelerometer a larger bandwidth does not show evident advantages over a lower bandwidth. For gyroscope, a larger bandwidth shows evident advantages over a smaller bandwidth. For instance, the subbands with 1 Hz bandwidth usually present low MI values. For magnetometer in Fig. 5(c), the upper block (energy ratio) provides more high-MI features than the lower block (absolute energy). This is in contrast to the other two modalities. For most class pairs, high MI values are observed in the frequency bands 0-15 Hz and 25-35 Hz. The bandwidth around 10 Hz seems to presents higher MI values than other bandwidths. This is consistent with the observations made in Fig. 4 that magnetometer provides more discriminablity between (S1/C5/B6) and (T7/S8).

#### 3) Quantile

Fig. 6 visualizes the MI values of various quantile features (family $\mathcal{Q}$) from the three modalities. The MI is computed between each feature and each of the 28 class pairs. Each subfigure contains 28 panesl corresponding to the 28 class pairs. The x- and y- axes denote the upper and lower bounds of a quantile range. Thus each cell with coordinate $(q_x, q_y)$ represents a quantile range value between $[q_y, q_x]$. The 9 specific quantile values, from 0 to 100, are listed in Table 9. A cell with the same coordinates, i.e. $q_x = q_y$, represent the quantile value $q_x$. The image in each panel resembles a lower-triangular area. Based on the MI values we can easily find out which quantile range has a higher discriminablity between the target classes.

For accelerometer in Fig. 6(a), the middle part of the triangular area in each panel tends to present higher MI values for most class pairs, e.g. the quantile range 25-75. For gyroscope in Fig. 6(b), the left part of the triangular area in each panel tends to present higher MI values for most class pairs, e.g. the quantile range 10-50. For magnetometer in Fig. 6(c), the right part of the triangular area in each panel tends to present higher MI values for most class pairs, e.g. the quantile range 0-100.

#### 4) Other time and frequency features

Fig. 7 visualizes the MI values of the time-frequency features from family $\mathcal{T}$+$\mathcal{F}$. The MI is computed between each feature and each of the 28 class pairs. Each subfigure contains 28 panels corresponding to the 28 class pairs. In each panel the indices 1-8 in the first column denote the time-domain features: mean, standard deviation, energy, mean crossing rate, kurtosis, skewness, auto-correlation value and offset.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



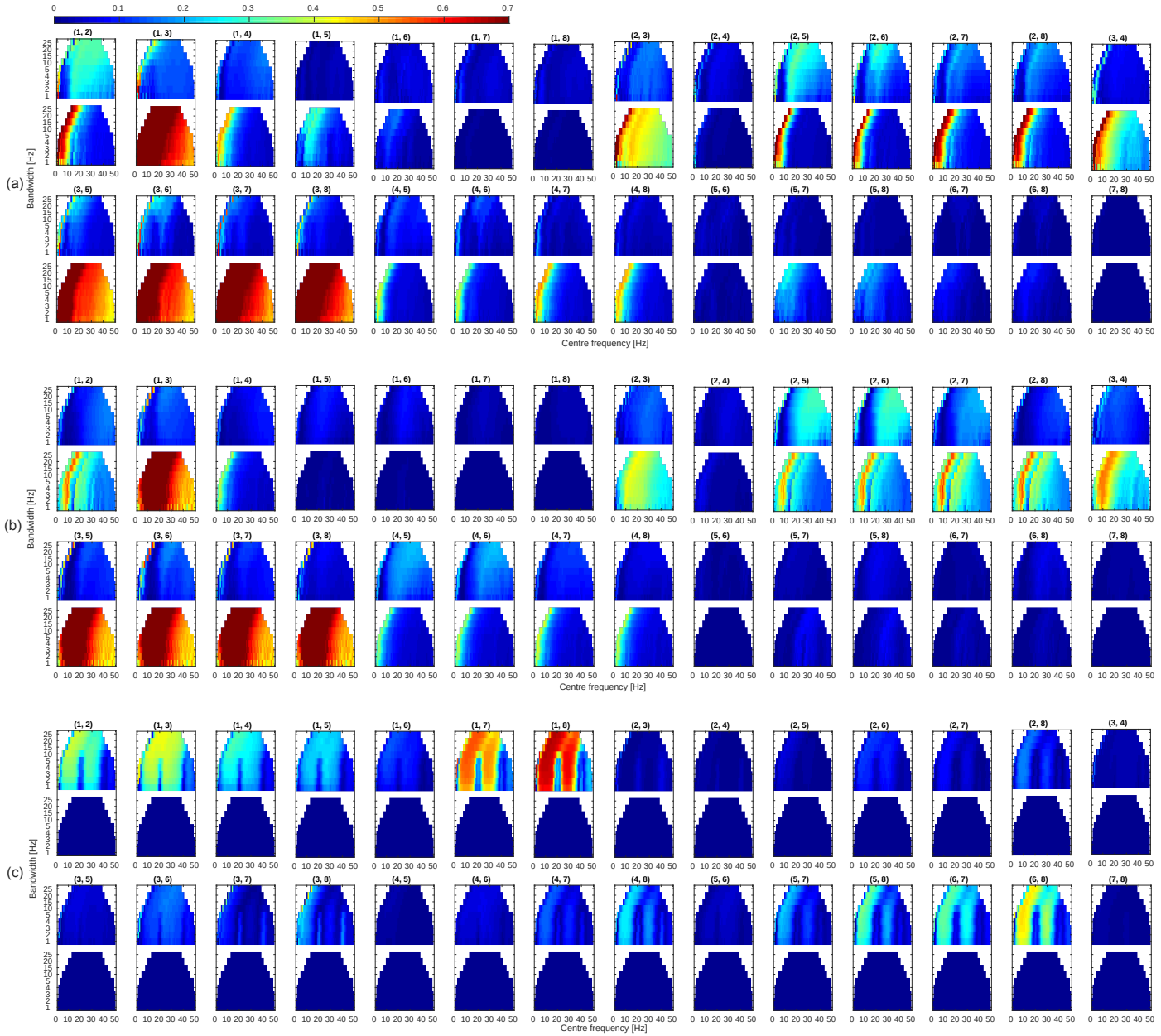


**FIGURE 5.** MI of subband features for (a) accelerometer; (b) gyroscope; (c) magnetometer. In each panel the upper block shows the MI of the energy-ratio features, and the lower block shows the MI of the absolute energy features. The x-axis denotes the center frequency while y-axis the bandwidth of the subband. Each subfigure contains 28 panels corresponding to 28 pair-wise combinations of the eight classes: S1-Still, W2-Walk, R3-Run, B4-Bike, C5-Car, B6-Bus, T7-Train, S8-Subway.

The indices 1-9 in the second column denote the frequency-domain features: DC, highest FFT value and frequency, ratio between the first and second peak, mean, standard deviation, kurtosis, skewness, and energy. It appears that all these features are important for one or more class pairs.

### C. FEATURE ANALYSIS BASED ON MRMR

The importance analysis in Sec. V-B relies only on the correlation between individual features and the target classes and does not consider the redundancy between the features. Since activity recognition usually uses multiple features, it is important to see which features will be selected after removing inter-feature redundancy.

MRMR is a well-known feature selection method which can select a set of features that has the maximum relevance with the target class and minimum redundancy between each other [58]. We thus employ MRMR to identify important features with least redundancy. Given the target classes $C$ and an initial set $\boldsymbol{F}$ with $n$ features, MRMR aims to find a subset $\boldsymbol{S} \subset \boldsymbol{F}$ with $k$ features that maximizes the mutual information between the features and the class $I(C;\boldsymbol{S})$ and minimize the mutual information between the features in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



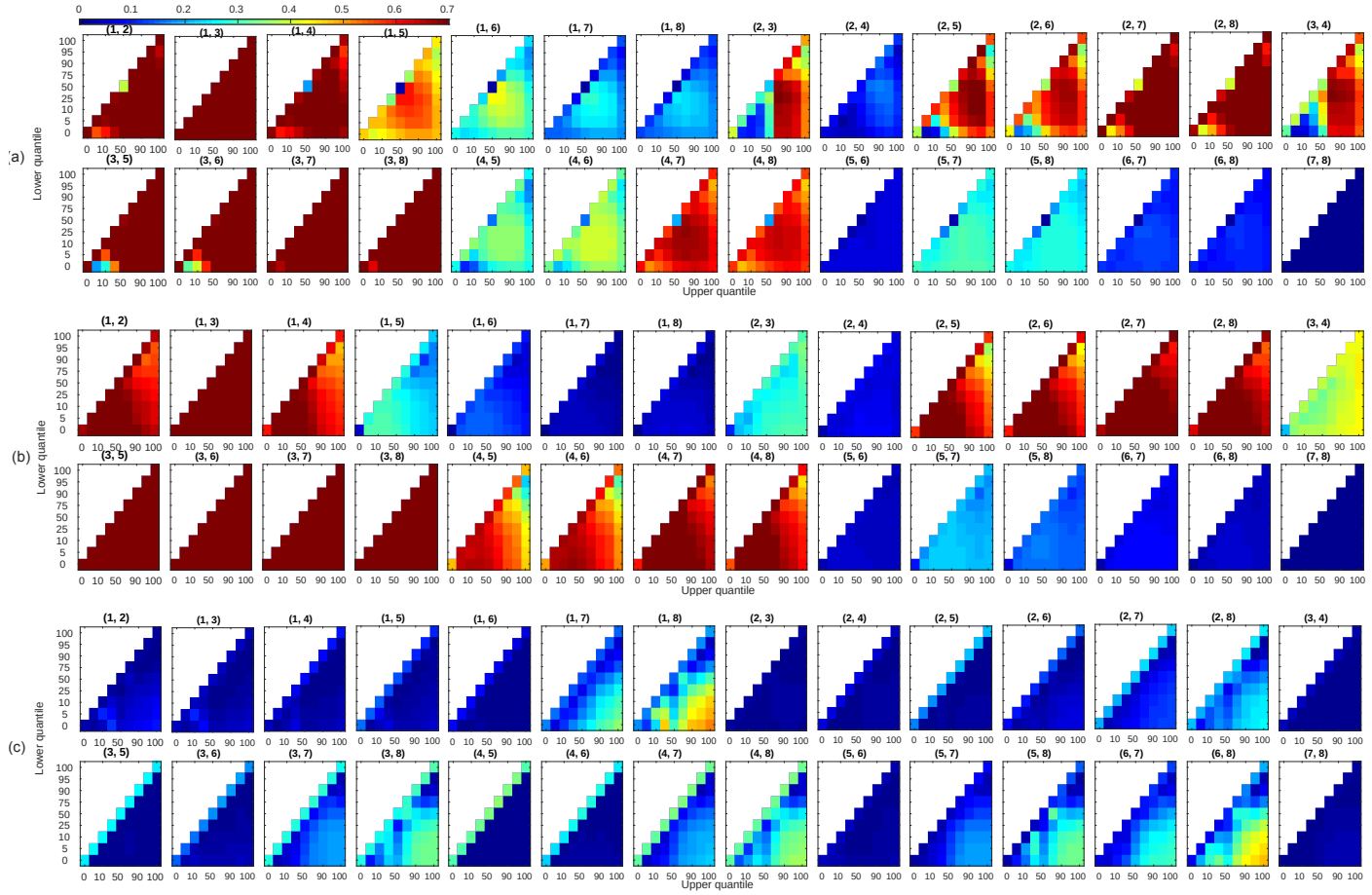


**FIGURE 6.** MI of quantile features for (a) accelerometer; (b) gyroscope; (c) magnetometer. The x-axis denotes the upper quantile while y-axis lower quantile. Each subfigure contains 28 panels corresponding to 28 pair-wise combinations of the eight classes: S1-Still, W2-Walk, R3-Run, B4-Bike, C5-Car, B6-Bus, T7-Train, S8-Subway.

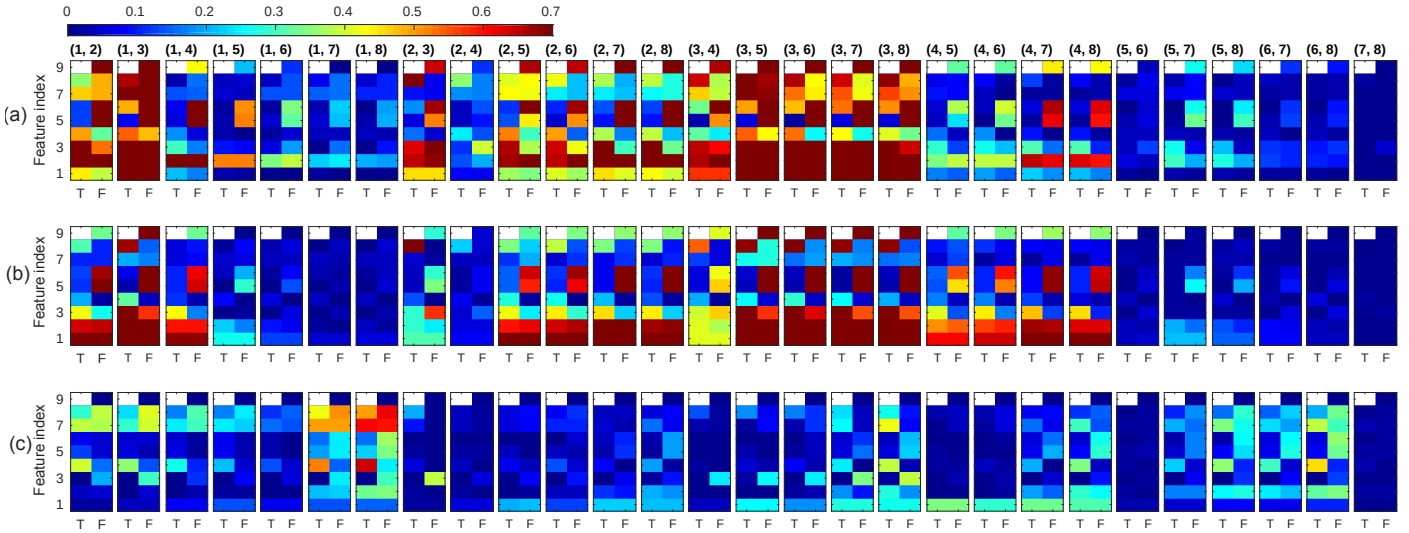


**FIGURE 7.** MI of time-domain (T) and frequency-domain (F) features for (a) accelerometer; (b) gyroscope; (c) magnetometer. Each subfigure contains 28 panels corresponding to 28 pair-wise combinations of the eight classes: S1-Still, W2-Walk, R3-Run, B4-Bike, C5-Car, B6-Bus, T7-Train, S8-Subway.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



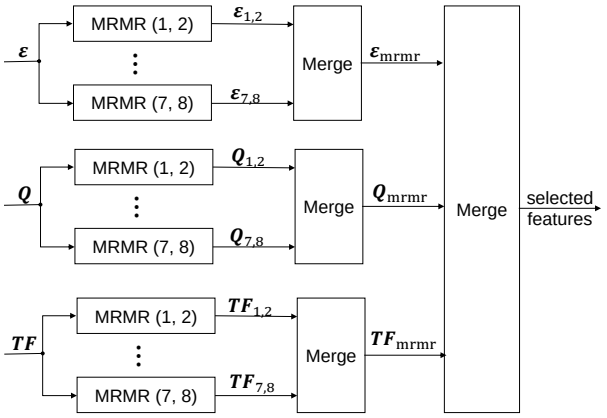


**FIGURE 8.** Block diagrams of the pair-wise MRMR feature selection method, which is applied separately to the three feature families: $\mathcal{E}$, $\mathcal{Q}$, $\mathcal{T}$+$\mathcal{F}$. A subset of features are selected for each of the 28 class pairs, and then merged together.

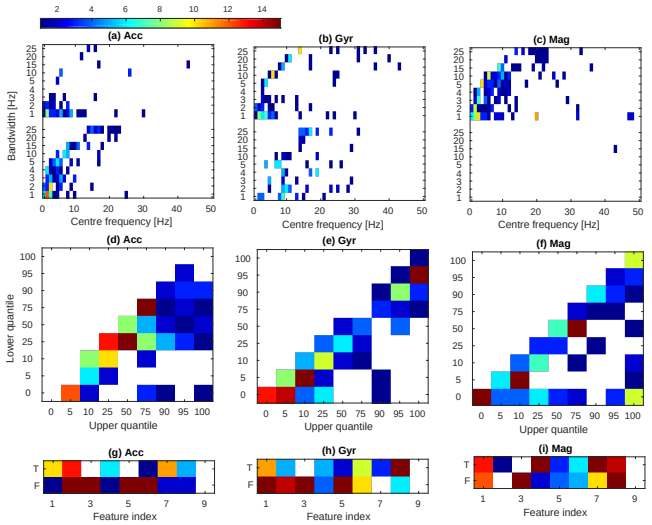


**FIGURE 9.** Merging the selected features from the 28 class pairs for each modality. The first row shows the selected subband features; the second row shows the selected quantile features; the third row shows the selected TF features. The color denotes the number of occurrence of each feature in the 28 class pairs.

the subset $I(\boldsymbol{f}_i;\boldsymbol{f}_j)$. An incremental search scheme is used which in each step selects a new feature that maximize the objective function $J(\boldsymbol{f}_i)$:

$$J(\boldsymbol{f}_i) = I(C;\boldsymbol{f}_i) - \frac{1}{|\boldsymbol{S}|}\sum_{\boldsymbol{f}_s\in\boldsymbol{S}}\frac{I(\boldsymbol{f}_s;\boldsymbol{f}_i)}{\min\{H(\boldsymbol{f}_i),H(\boldsymbol{f}_s)\}}, \quad (7)$$

where $H(\boldsymbol{f}_i)$ denotes the entropy of the feature $\boldsymbol{f}_i$, and $\boldsymbol{f}_s$ denotes a feature in the subset $\boldsymbol{S}$. The normalization in the second term of (7) aims to limit the MI within the range $[0,1]$ in order to prevent over-weighting nonredundant features [60].

As shown in Sec. V-B, each feature presents different MI values for different class pairs, and consequently each class pair leads to a different optimal set of features according to the MRMR criterion. To avoid removing features that are potentially useful, we perform feature selection per class pair and per modality by applying MRMR independently to each of the three feature families: $\mathcal{E}$, $\mathcal{Q}$, $\mathcal{T}$+$\mathcal{F}$. Fig. 8 depicts the block diagrams of the pair-wise MRMR feature selection method.

For each modality, we select 10 features from $\mathcal{E}$ for each class pair and then combine selected features from the 28 class pairs together. This procedure is repeated for $\mathcal{Q}$ (5 features per class pair) and $\mathcal{T}$+$\mathcal{F}$ (5 features per class pair). Fig. 9 illustrates the selected features from the families $\mathcal{E}$, $\mathcal{Q}$, and $\mathcal{T}$+$\mathcal{F}$ for the three modalities. It may happen that some class pairs lead to the selection of the same feature. We thus use color to indicate how often a feature is selected, which can range from 'never' up to a feature being selected 28 times, i.e. once for each of the 28 class pairs. The more frequently selected, the more important a feature is. A summary on the selection result is given below.

For accelerometer the MRMR algorithm produces 147 features including 104 subband features ($\mathcal{E}$), 29 quantile features ($\mathcal{Q}$) and 14 time-frequency features ($\mathcal{T}$+$\mathcal{F}$). The most selected subband features (Fig. 9(a)) tend to have a center frequency between 0 and 5 Hz and a bandwidth between 1 and 5 Hz. These features appear in both upper block (energy ratio) and lower block (absolute energy) of Fig. 9(a). The most selected quantile features (Fig. 9(d)) appear on the left side of the triangular area with a narrow interval between lower and upper quantiles. For TF features in Fig. 9(g), most features are selected except two time-domain features (energy and kurtosis) and one frequency-domain feature (energy).

**TABLE 10.** Five most frequently reoccurring subband, quantile and TF features in the 28 class pairs for each modality. Key: $\omega_c$ - center frequency; $\omega_b$ - bandwidth.

| | Accelerometer | Gyroscope | Magnetometer |
|---|---|---|---|
| Subband | energy: $\omega_c$=2Hz, $\omega_b$=1Hz<br>energy: $\omega_c$=12Hz, $\omega_b$=2Hz<br>ratio: $\omega_c$=2Hz, $\omega_b$=1Hz<br>energy: $\omega_c$=3Hz, $\omega_b$=1Hz<br>energy: $\omega_c$=1Hz, $\omega_b$=1Hz | ratio: $\omega_c$=9Hz, $\omega_b$=10Hz<br>ratio: $\omega_c$=14Hz, $\omega_b$=25Hz<br>ratio: $\omega_c$=2Hz, $\omega_b$=1Hz<br>energy: $\omega_c$=9Hz, $\omega_b$=1Hz<br>ratio: $\omega_c$=3Hz, $\omega_b$=1Hz | ratio: $\omega_c$=2Hz, $\omega_b$=1Hz<br>ratio: fc=1Hz, $\omega_b$=1Hz<br>ratio: $\omega_c$=20Hz, $\omega_b$=1Hz<br>ratio: $\omega_c$=3Hz, $\omega_b$=1Hz<br>ratio: $\omega_c$=4Hz, $\omega_b$=5Hz |
| Quantile | Q75<br>Q25-Q50<br>Q25<br>Q0-Q5<br>Q10-Q25 | Q5-Q10<br>Q95-Q100<br>Q0-Q5<br>Q0<br>Q10-Q25 | Q0<br>Q5-Q10<br>Q50-Q75<br>Q100<br>Q0-Q100 |
| TF | highest FFT value<br>highest FFT frequency<br>mean of FFT<br>std of FFT<br>std of samples | mean of FFT<br>highest autocorr index<br>highest FFT frequency<br>DC of FFT<br>highst FFT value | mean crossing rate<br>highest autocorr value<br>highest FFT frequency<br>skewness of FFT<br>highest autocorr value |

For gyroscope the MRMR algorithm produces 150 features including 108 subband features ($\mathcal{E}$), 28 quantile features ($\mathcal{Q}$) and 14 time-frequency features ($\mathcal{T}$+$\mathcal{F}$). The most selected subband features (Fig. 9(b)) tend to distribute sparsely at subbands with a center frequency between 0 and 30 Hz, and a bandwidth between 1 and 25 Hz. These features appear in both upper block (energy ratio) and lower block (absolute energy) in Fig. 9(b). The most selected quantile features (Fig. 9(e)) tend to appear at the left side of the triangular shape, with a narrow interval between lower and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



upper quantiles. For TF features in Fig. 9(e), most features are selected except one time-domain features (energy) and two frequency-domain feature (kurtosis and energy).

For magnetometer, the MRMR algorithm produces 148 features including 104 subband features ($\mathcal{E}$), 30 quantile features ($\mathcal{Q}$) and 14 time-frequency features ($\mathcal{T}$+$\mathcal{F}$). The most selected subband features (Fig. 9(c)) appear in the upper block (energy ratio) and very few appear in the lower block (absolute energy). These features tend to distribute densely at subbands with a center frequency between 0 and 15 Hz and a bandwidth between 1 and 10 Hz, and also tend to distribute at subbands with a center frequency between 20 and 30 Hz and a bandwidth between 20 and 25 Hz. The most selected quantile features (9(f)) tend to appear at the left side of the triangular shape, with a narrow interval between lower and upper quantiles. However, a feature covering the full range between quantile 0 and quantile 100 is also selected for multiple times. For TF features in Fig. 9(i), most features are selected except one time-domain features (energy) and two frequency-domain feature (highest FFT value and energy).

Finally, Table 10 lists the five most frequently reoccurring features in $\mathcal{E}$, $\mathcal{Q}$, $\mathcal{T}$+$\mathcal{F}$, respectively in each modality.

Note that while the proposed MRMR-based feature analysis procedure is computationally expensive, this computation only occurs when the system is developed, i.e. in the training stage. At run-time, in a deployed system, only the selected features need to be computed (i.e. MRMR needs not be run in a production system, only during development). This reduces the computation significantly in the deployed system as less features are computed and used for the classification.

To summarize, the significance analysis in Sec. V-B and Sec. V-C gives us an idea on which feature provides crucial information for a specific recognition task. We can use the features selected in this section as a starting point to establish the baseline performance of the defined recognition tasks.

## VI. BASELINE PERFORMANCE

### A. PROCESSING PIPELINE

Fig. 10 illustrates the processing pipeline for establishing baseline performance for the recommended recognition tasks using the SHL dataset.

We compute the recognition performance for each recognition task which is defined as a combination of leave-one-out scheme, an evaluation scenario, and a group of modalities in Table 7. We first divide the entire dataset into training and testing folds according to the leave-one-out evaluation strategy indicated in Table 7. For the training dataset, we use a sliding window with a length of 5.12 seconds and 2.56-second overlap to segment the sensor data into frames and in each frame we extract a set of features $\{f_e\}$ identified in Sec. V-C, including 147 accelerometer features, 150 gyroscope features and 148 magnetometer features (Fig. 9). For each of the 12 evaluation scenarios, we apply MRMR to select 50 features independently for each of the three modalities: accelerometer, gyroscope and magnetometer, and compute two features for the GPS modality: mean

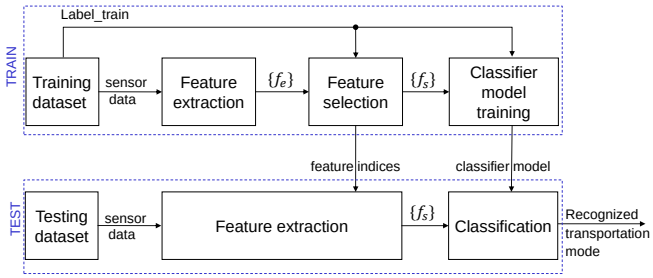


**FIGURE 10.** The processing pipeline using the SHL dataset, which is divided into the training and testing datasets according to the leave-one-out strategy. The training dataset is used for feature selection and classifier model training (top block). The testing dataset is used for performance evaluation (bottom block).

speed and mean acceleration. The speed and acceleration is estimated based on the change of GPS coordinates (latitude and longitude) over time with the Matlab Mapping Toolbox. For each group of modalities, we combine all the features computed on each constituent modality in a single feature vector $\{f_s\}$. For instance, the feature vector of the modality group AGM consists of 150 elements. The resulting feature vector and associated class label corresponding to each frame of data in the train set are used to train the classifier model. The testing dataset comprises all the data frames in the left-out fold of the cross-validation. Based on the indices of the features selected in the training stage, we compute the same set of features $\{f_s\}$ and feed them to the trained classifier model to recognize the transportation mode in each frame.

We employ a decision tree as a baseline classifier due to its popularity in transportation mode recognition (e.g. 18 out of 34 related work). We implemented the recognition system using Matlab's built-in function 'fitctree'. We use the default parameter for this function except setting the parameter 'MinParentSize' (the minimum number of observations per branch node in the tree) to $\frac{10000}{C}$, where $C$ is the number of the classes for a specific recognition task, and setting the parameter 'MinLeafSize' (the minimum number of observations per leaf node in the tree) as $\frac{\text{MinParentSize}}{5}$. We use large values for these two parameters to prevent overfitting in the training stage.

As already discussed in Sec. IV, the evaluation of the groups of modalities A, AG and AGM will be made on Dataset-E and Dataset-IG, respectively, and the evaluation of P, AP and AGMP will be made on Dataset-IG.

### B. RESULTS

Table 11 reports in detail the baseline performance, in terms of recognition accuracy and F1 score, of the 396 recognition tasks, consisting of 12 evaluation scenarios, 11 leave-one-out cross-validations (three users, four positions and four periods), and three groups of modalities (A, AG and AGM) obtained using Dataset-E. Table 12 reports the baseline performance of the 729 recognition tasks with six groups of modalities (A, AG, AGM, P, AP and AGMP) obtained using Dataset-IG.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



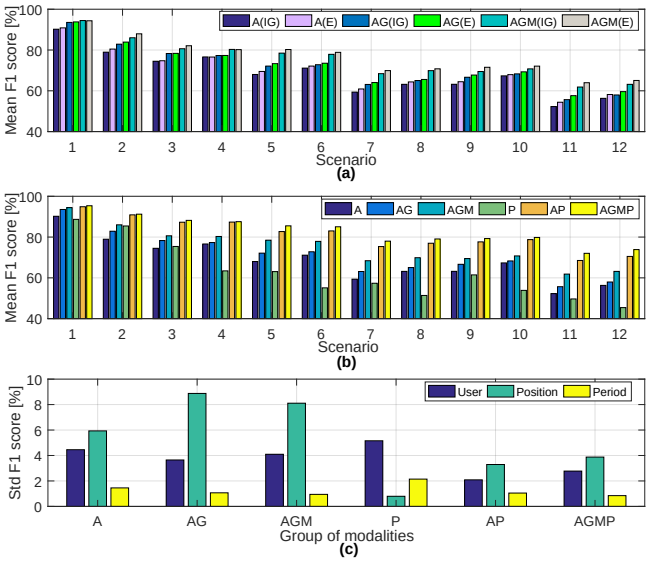


**FIGURE 11.** Visualization of the F1 score results. (a) The mean F1 score for each scenario obtained by the modalities A, AG and AGM, and with Dataset-E and Dataset-IG. (b) The mean F1 score for each scenario obtained by the modalities A, AG, AGM, P, AP and AGMP, and with Dataset-IG. (c) The standard deviation of the F1 score across users, positions and periods for each group of modalities (Dataset-IG).

To investigate the influence of different modalities on the recognition performance, we average the F1 scores on all the 11 cross-validation cases for each recognition scenario and each group of modalities. Fig. 11(a) depicts the mean F1 score for the 12 recognition scenarios and three groups of modalities (A, AG and AGM), obtained using Dataset-E and Dataset-IG, respectively. Fig. 11(b) depicts the mean F1 score for the 12 scenarios and six modality groups, obtained using Dataset-IG. Regardless of their different amount of data, Dataset-E and Dataset-IG achieve very similar F1 scores for all groups of modalities and recognition scenarios. Meanwhile, the F1 score by Dataset-E is slightly higher than that by Dataset-IG, possibly because the former one has a more balanced data between classes. For each recognition scenario, using more modalities appears to always increase the recognition performance. Specifically, the following observations can be made.

- The combination of accelerometer and gyroscope (AG) tends to improve the recognition performance over that obtained with an accelerometer alone (A) slightly.
- Including the magnetometer (AGM) tends to improve the recognition performance much more significantly. The pronounced improvement by combining accelerometer and magnetometer is due to the complementarity between the two, i.e. one is based on the motion of the device while the other is based on the ambient magnetic field around the device. As shown in Fig. 4, the magnetometer tends to provide more features with high MI values for class pairs where accelerometer and gyroscope provide very few features with high MI values.
- The GPS modality alone, with only two features, does not provide sufficient discriminablility between the target classes. However, combining GPS and accelerometer (AP) tends to improve the recognition performance significantly over using either modality alone (A or P). The combination of GPS and accelerometer (AP) outperforms the combination of three inertial sensors (AGM). The combination of all the four modalities (AGMP) only improves the recognition performance slightly over AP.

We use the standard deviation of F1 score to investigate the influence of user, position and temporal variation on the recognition performance. For user variation, we compute the standard deviation of F1 score across three users per recognition scenario and per group of modalities, and then average the standard deviation values across the 12 recognition scenarios for each group of modalities. We repeat the same procedure for position variation (with four positions) and temporal variation (with four periods). All the results are obtained using Dataset-IG. Fig. 11(c) depicts the standard deviation for the three variations (user, position, and period) and six groups of modalities, where a smaller standard deviation implies more robustness of recognition system to the variation. The following observations can be made.

- When using inertial sensors (A, AG, AGM), the position variation tends to introduce the largest standard deviation among the three, because human engages with the recording device differently depending on the wearing position. It can be observed in both Table 11 and Table 12 the recognition performance at the four positions can be ranked as Hand $>$ Torso $>$ Hips $>$ Bag.
- When using both inertial and GPS sensors the standard deviation of position variation is reduced significantly. This demonstrates that GPS can increase the robustness of the recognition system to position variation, because GPS information does not vary much with wearing positions. When using inertial sensors only, the user variation has the second largest standard deviation because each user has a different behaviour style during the travel.
- When using GPS alone, the user variation appears to have the largest standard deviation of the recognition performance. This is possibly because each user has a different speed when performing walking, running, biking and driving activities. The temporal variation tends to have the smaller standard deviation of the recognition performance across all the five groups of modalities (except P - GPS alone).

Fig. 12 lists the confusion matrices for Scenario 12 (the most difficult scenario with eight classes) evaluated on Period 3 (leave-one-period-out cross-validation). The first row shows the results for the three groups of modalities (A, AG and AGM) obtained with Dataset-E. The second and third

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



rows show the results for the six groups of modalities (A, AG, AGM, P, AP and AGMP) obtained with Dataset-IG. From the confusion matrices, we can draw similar conclusions as we did from Fig. 11. As shown in the first and the second rows of Fig. 12, Dataset-E and Dataset-IG achieve a similar recognition accuracy for A, AG and AGM, whereas Dataset-E achieves a slightly higher F1 score than Dataset-IG. This is because that Dataset-E has more balanced data between the eight classes, as supported by the recognition result for the class S8 - Subway, where Dataset-E achieves a much higher recognition accuracy (e.g. 53.8% vs 30.3% for AGM in the confusion matrix).

From the confusion matrices in the second and third rows we can clearly see how the recognition performance is improved by using more modalities. Specifically, the following observations can be made.

- When using accelerometer (A) alone, the classifier can recognize Still, Walk and Run robustly, but presents significant ambiguities between Car and Bus, and between Train and Subway, and certain ambiguities between Still and Train/Subway, and relatively low recognition rate of Bike. Car and Bus may have similar sensor vibration intensity, thus leading to larger confusion between each other; so does the pair Train and Subway. Bike may be mis-recognized as Walk, Bus or Car, each with a probability of around 7%.
- When combining accelerometer and gyroscope (AG), the classifier can better recognize the Bike, whose recognition accuracy is improved from 76.5% to 84.5%. This is possibly because biking activities involves more rotational behaviours, e.g. turning often the handlebar of the bicycle when cycling.
- When magnetometer is included to AG, denoted AGM, the recognition accuracy of Subway is improved notably from 32.4% to 53.8%. The ambiguity between Still and Train/Subway is also reduced significantly.
- When using GPS alone, the classifier presents a very low recognition accuracy for Run (7%) and tends to misclassify it as Bike and Walk. This is due to the fact that the running speed of some of the subjects may not have been significantly faster than walking, or in a similar range to leisurely cycling. The classifier also presents a very low recognition accuracy for Subway (0.7%) and tends to misclassify it as Car and Bus. This is linked to the speed of the vehicles: a subway is 40-60 km per hour, which is similar to bus, and often to car in cities.
- When combining GPS and accelerometer (AG), the recognition accuracy for each class is improved remarkably in comparison to using accelerometer alone (A). In particularly, the recognition accuracy of Car and Bus has each been improved from 45% to around 70%. Train can be better recognized with the accuracy improved from 51% to 67%.
- Comparing AGM and AGMP, the latter one improves the recognition accuracy of Bus, Car and Train remarkably with each above 10%, but achieves a decreased recognition rate of Subway. This is possibly because Subway does not have sufficient GPS data available, thus leading to a biased classification result. Interestingly, the availability of GPS does show a strong indication of Still (inside) or Subway. This fact could be be further exploited to improve the recognition performance.

## VII. DISCUSSION

We recommend 792 recognition tasks as a combination of 12 recognition scenarios, six groups of modalities, and three leave-one-out cross-validation evaluation criteria to be used by the research community for a standardized comparison. These recognition tasks are defined based on the SHL dataset and constitute a superset covering the majority recognition tasks considered in the literature, except some transportation activities not included in the SHL dataset. We suggest to use the naming scheme "Task-Scenario-Crossvalidation-Modality" when performing a specific evaluation task using the SHL dataset. Here 'Scenario' can be 'O1-O12'; 'Crossvalidation' can be 'UX', 'PX', and 'TX' denoting user-independent, position-independent, and time-invariant evaluation with folder 'X' out; 'Modality' can be 'A', 'AG', 'AGM', 'P', 'AP' and 'AGMP' (see Table 7). For instance, "Task-O12-U1-A" denotes the leave-User1-out evaluation on Scenario 12 using the accelerometer modality, while 'Task-O2-P2-AP" denotes the leave-Torso-out evaluation on Scenario 2 using the accelerometer and GPS modalities. With this naming scheme we can easily associate a specific recognition task in the related work with the one defined in this paper. For instance, related work [49] addressed Scenario 4, with an 'user-independent', 'position-independent' and 'time-invariant' evaluation using the group of modalities 'AP'. The authors of [49] would be able to apply their algorithms to SHL dataset and compare with baseline results reported in this paper (e.g. Table 12). In case that the average performance of cross-validation is reported, we recommend to use the name 'Task-O12-P-AP' to represent the average position-indepedent cross-validation performance for Scenario 12 using the accelerometer and GPS modalities.

In this paper we mainly aim to establish a standard performance evaluation framework rather than pursuing the maximum recognition performance. The recognition pipeline presented in this paper is a baseline implementation, which aims to provide reference results to enable reproducible comparison. For this reason, we employ a well understood classifier, the decision tree, in our pipeline. In fact, the recognition performance is affected by several aspects including the features, classifiers and the recognition tasks. All the observations and conclusions made in this paper are confined to the baseline implementation. However, all the feature analysis results presented in Sec. V are classifier-agnostic. In particular, our identification of relevant

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



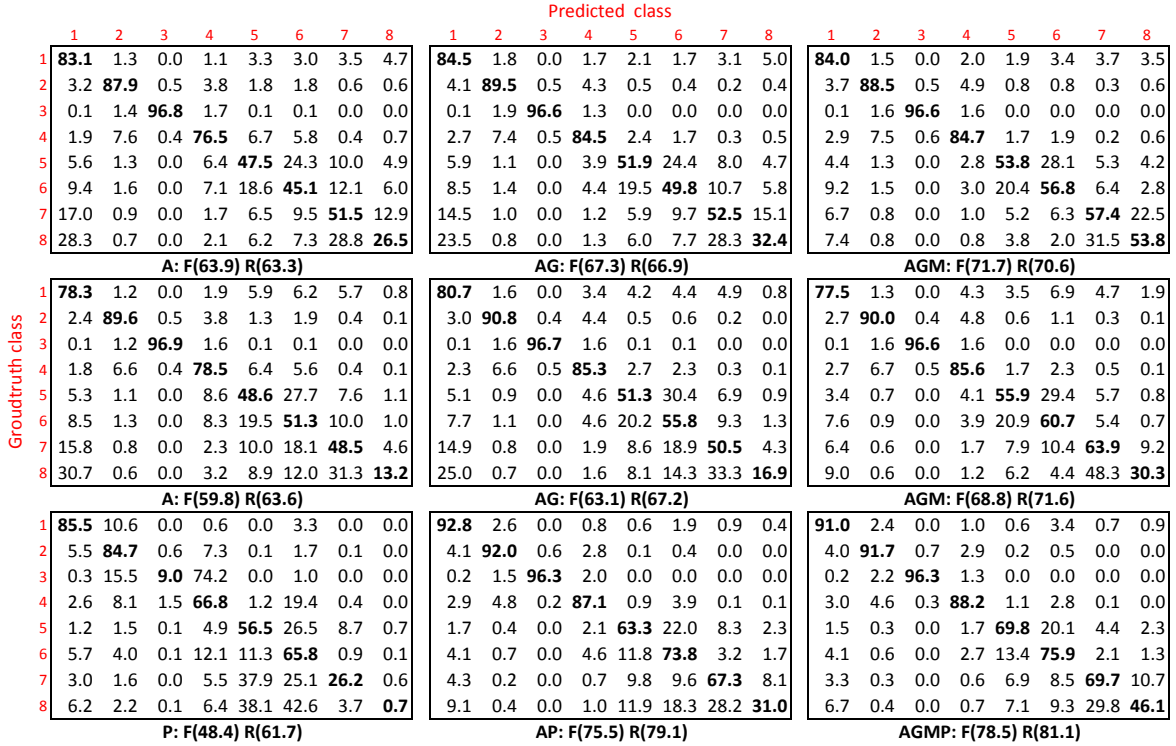


**FIGURE 12.** Confusion matrices for Scenario 12 evaluated on Period 3 (time-invariant cross-validation). The first row is obtained using Dataset-E. The second and third rows are obtained using Dataset-IG. Eight classes: S1-Still, W2-Walk, R3-Run, B4-Bike, C5-Car, B6-Bus, T7-Train, S8-Subway.

**TABLE 11.** F1 score (F) and recognition accuracy (R) for each recognition task obtained using Dataset-E (the entire dataset).

| | | User1 | | User2 | | User3 | | Avg | | Pos1 | | Pos2 | | Pos3 | | Pos4 | | Avg | | Fold1 | | Fold2 | | Fold3 | | Fold4 | | Avg | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Scenario | F | R | F | R | F | R | F | R | F | R | F | R | F | R | F | R | F | R | F | R | F | R | F | R | F | R | F | R |
| A | 1 | 91.8 | 92.9 | 89.2 | 91.2 | 84.9 | 88.0 | **88.6** | **90.7** | 86.3 | 88.5 | 93.6 | 94.7 | 91.7 | 93.1 | 92.9 | 94.1 | **91.1** | **92.6** | 92.3 | 93.7 | 93.2 | 94.3 | 93.3 | 94.1 | 92.4 | 94.1 | **92.8** | **94.1** |
| | 2 | 79.4 | 83.9 | 77.2 | 80.6 | 76.9 | 78.3 | **77.8** | **80.9** | 74.8 | 78.3 | 83.9 | 86.5 | 75.7 | 80.5 | 87.2 | 88.5 | **80.4** | **83.4** | 83.8 | 85.8 | 83.9 | 86.9 | 82.6 | 85.6 | 82.4 | 84.6 | **83.2** | **85.7** |
| | 3 | 66.2 | 76.6 | 72.9 | 78.9 | 68.1 | 75.6 | **69.1** | **77.0** | 66.1 | 73.0 | 70.4 | 77.7 | 75.6 | 79.7 | 84.4 | 87.4 | **74.1** | **79.4** | 81.5 | 84.4 | 81.8 | 85.3 | 81.3 | 83.9 | 79.6 | 83.2 | **81.0** | **84.2** |
| | 4 | 72.0 | 77.2 | 73.1 | 79.0 | 65.1 | 74.4 | **70.1** | **76.9** | 70.1 | 72.5 | 74.4 | 78.5 | 76.2 | 77.7 | 85.3 | 86.4 | **76.5** | **78.8** | 82.4 | 83.7 | 84.3 | 85.1 | 83.8 | 83.5 | 82.0 | 82.9 | **83.1** | **83.8** |
| | 5 | 62.7 | 65.8 | 66.4 | 66.5 | 63.3 | 66.5 | **64.1** | **66.3** | 60.4 | 61.4 | 67.7 | 68.4 | 69.5 | 68.9 | 79.3 | 78.9 | **69.2** | **69.4** | 75.8 | 75.2 | 76.6 | 76.4 | 75.4 | 74.7 | 73.0 | 72.5 | **75.2** | **74.7** |
| | 6 | 67.5 | 66.2 | 67.5 | 66.3 | 62.2 | 66.3 | **65.7** | **66.3** | 64.8 | 60.7 | 73.6 | 71.0 | 72.4 | 68.2 | 81.4 | 78.5 | **73.1** | **69.6** | 76.6 | 73.6 | 79.5 | 76.1 | 78.5 | 74.5 | 75.8 | 71.7 | **77.6** | **74.0** |
| | 7 | 55.2 | 58.4 | 58.2 | 60.6 | 52.7 | 63.6 | **55.3** | **60.9** | 51.1 | 53.2 | 61.5 | 64.1 | 60.8 | 63.1 | 68.4 | 69.9 | **60.5** | **62.6** | 67.5 | 67.9 | 68.8 | 69.5 | 66.8 | 68.8 | 64.6 | 69.5 | **66.9** | **68.9** |
| | 8 | 59.9 | 58.3 | 60.6 | 60.5 | 52.7 | 62.8 | **57.7** | **60.5** | 57.2 | 54.6 | 65.1 | 63.9 | 64.5 | 62.5 | 72.2 | 70.2 | **64.8** | **62.8** | 70.3 | 67.7 | 72.4 | 69.2 | 70.7 | 68.7 | 69.0 | 69.6 | **70.6** | **68.8** |
| | 9 | 56.6 | 63.6 | 62.3 | 66.3 | 54.4 | 71.2 | **57.8** | **67.0** | 54.8 | 60.2 | 63.5 | 69.2 | 64.5 | 68.7 | 73.7 | 75.5 | **64.1** | **68.4** | 72.8 | 73.5 | 72.4 | 73.7 | 71.4 | 74.0 | 69.1 | 76.4 | **71.4** | **74.4** |
| | 10 | 62.5 | 63.6 | 64.8 | 67.1 | 54.4 | 70.4 | **60.6** | **67.0** | 62.0 | 61.4 | 65.3 | 66.9 | 68.2 | 68.4 | 77.1 | 75.5 | **68.1** | **68.1** | 74.9 | 72.7 | 76.6 | 74.0 | 75.1 | 73.8 | 73.7 | 76.7 | **75.1** | **74.3** |
| | 11 | 50.0 | 53.5 | 52.0 | 55.6 | 47.0 | 56.3 | **49.7** | **55.1** | 46.2 | 48.8 | 54.4 | 57.6 | 54.6 | 58.0 | 60.3 | 63.9 | **53.9** | **57.1** | 59.9 | 64.2 | 61.7 | 64.8 | 60.0 | 64.0 | 56.9 | 60.5 | **59.6** | **63.4** |
| | 12 | 54.9 | 53.5 | 54.7 | 55.3 | 47.1 | 55.1 | **52.3** | **54.6** | 52.5 | 50.3 | 59.2 | 58.2 | 58.1 | 57.0 | 65.5 | 64.4 | **58.8** | **57.5** | 63.0 | 63.7 | 65.6 | 64.5 | 63.9 | 63.3 | 61.5 | 60.4 | **63.5** | **63.0** |
| AG | 1 | 94.2 | 95.1 | 92.9 | 94.1 | 91.1 | 92.5 | **92.7** | **93.9** | 88.4 | 89.7 | 95.6 | 96.4 | 94.8 | 95.7 | 94.8 | 95.7 | **93.4** | **94.4** | 95.3 | 96.1 | 95.3 | 96.1 | 95.2 | 95.7 | 95.1 | 96.1 | **95.2** | **96.0** |
| | 2 | 82.5 | 86.5 | 80.6 | 85.0 | 84.3 | 85.3 | **82.5** | **85.6** | 75.5 | 79.6 | 86.8 | 89.4 | 81.9 | 85.7 | 89.4 | 90.9 | **83.4** | **86.4** | 87.0 | 89.1 | 85.7 | 89.1 | 85.1 | 88.3 | 85.4 | 87.6 | **85.8** | **88.5** |
| | 3 | 70.9 | 80.4 | 77.4 | 82.1 | 75.9 | 80.9 | **74.7** | **81.2** | 65.3 | 72.0 | 71.7 | 79.6 | 79.5 | 83.1 | 86.8 | 89.2 | **75.8** | **81.0** | 86.1 | 88.0 | 84.4 | 87.6 | 83.9 | 86.4 | 83.7 | 86.1 | **84.5** | **87.0** |
| | 4 | 72.8 | 79.4 | 74.6 | 81.9 | 67.1 | 78.9 | **71.5** | **80.1** | 58.0 | 70.6 | 72.0 | 78.5 | 81.0 | 82.5 | 87.9 | 89.2 | **74.7** | **80.2** | 85.5 | 87.0 | 86.7 | 87.6 | 85.9 | 86.2 | 85.1 | 85.8 | **85.8** | **86.6** |
| | 5 | 66.8 | 70.5 | 70.7 | 70.4 | 70.9 | 72.8 | **69.5** | **71.2** | 60.9 | 61.3 | 68.7 | 70.4 | 74.8 | 74.1 | 81.9 | 81.5 | **71.6** | **71.8** | 80.1 | 79.6 | 79.7 | 79.6 | 78.5 | 77.9 | 77.4 | 76.2 | **78.9** | **78.3** |
| | 6 | 69.1 | 69.1 | 71.2 | 70.8 | 66.5 | 71.4 | **68.9** | **70.4** | 54.2 | 59.1 | 69.1 | 68.5 | 76.5 | 73.1 | 83.5 | 81.1 | **70.8** | **70.4** | 80.4 | 78.2 | 82.4 | 79.4 | 81.1 | 77.6 | 79.5 | 75.6 | **80.8** | **77.7** |
| | 7 | 57.4 | 60.5 | 62.0 | 63.7 | 57.2 | 68.6 | **58.9** | **64.3** | 53.5 | 55.1 | 59.8 | 63.3 | 65.2 | 67.4 | 71.5 | 72.6 | **62.5** | **64.6** | 71.8 | 71.9 | 72.0 | 72.4 | 70.4 | 72.0 | 68.9 | 73.2 | **70.8** | **72.4** |
| | 8 | 61.0 | 59.7 | 62.8 | 63.5 | 54.6 | 66.5 | **59.5** | **63.2** | 49.3 | 52.6 | 61.2 | 60.9 | 68.4 | 66.6 | 74.4 | 72.6 | **63.4** | **63.2** | 74.1 | 71.6 | 75.2 | 72.2 | 73.5 | 71.7 | 72.6 | 73.2 | **73.9** | **72.2** |
| | 9 | 60.2 | 66.1 | 65.7 | 69.2 | 60.4 | 76.2 | **62.1** | **70.5** | 56.3 | 60.8 | 60.8 | 68.0 | 69.6 | 73.2 | 76.0 | 77.6 | **65.7** | **69.9** | 77.3 | 77.1 | 75.8 | 76.5 | 74.9 | 76.9 | 73.7 | 79.8 | **75.4** | **77.6** |
| | 10 | 63.4 | 65.3 | 66.8 | 69.8 | 59.5 | 75.9 | **63.2** | **70.3** | 51.6 | 59.5 | 62.7 | 66.0 | 72.8 | 72.8 | 78.9 | 78.0 | **66.5** | **69.1** | 78.5 | 76.2 | 79.3 | 76.5 | 77.9 | 76.5 | 77.0 | 79.6 | **78.2** | **77.2** |
| | 11 | 53.1 | 56.4 | 55.9 | 58.7 | 51.5 | 61.2 | **53.5** | **58.8** | 45.6 | 48.5 | 54.6 | 57.5 | 58.9 | 61.5 | 63.6 | 66.9 | **55.7** | **58.6** | 64.3 | 68.3 | 65.3 | 68.1 | 63.5 | 67.1 | 61.6 | 64.5 | **63.7** | **67.0** |
| | 12 | 55.5 | 54.5 | 57.3 | 58.7 | 50.3 | 59.5 | **54.4** | **57.5** | 44.7 | 48.2 | 55.4 | 54.2 | 62.1 | 60.6 | 67.6 | 67.4 | **57.5** | **57.6** | 67.0 | 67.7 | 68.7 | 67.6 | 67.3 | 66.9 | 64.9 | 63.8 | **67.0** | **66.5** |
| AGM | 1 | 95.2 | 96.0 | 93.2 | 94.4 | 91.3 | 92.7 | **93.2** | **94.4** | 90.5 | 91.6 | 95.9 | 96.5 | 95.5 | 96.3 | 95.2 | 96.0 | **94.3** | **95.1** | 95.5 | 96.2 | 95.7 | 96.4 | 95.4 | 95.9 | 95.5 | 96.4 | **95.5** | **96.2** |
| | 2 | 88.5 | 91.1 | 84.8 | 88.0 | 87.6 | 88.6 | **86.9** | **89.2** | 81.5 | 84.9 | 89.5 | 91.7 | 88.5 | 90.5 | 90.8 | 92.3 | **87.6** | **89.8** | 90.1 | 91.6 | 89.1 | 91.6 | 88.4 | 90.9 | 89.7 | 91.5 | **89.3** | **91.4** |
| | 3 | 75.0 | 84.2 | 81.9 | 86.1 | 80.1 | 84.6 | **79.0** | **85.0** | 71.1 | 77.5 | 77.0 | 84.4 | 85.4 | 88.7 | 87.4 | 90.3 | **80.2** | **85.2** | 87.8 | 90.0 | 86.7 | 89.9 | 86.5 | 89.0 | 87.2 | 90.1 | **87.1** | **89.7** |
| | 4 | 76.3 | 83.7 | 79.7 | 86.6 | 70.3 | 82.7 | **75.4** | **84.3** | 62.0 | 75.8 | 74.4 | 81.7 | 84.9 | 87.4 | 88.2 | 90.0 | **77.4** | **83.7** | 87.3 | 89.3 | 88.2 | 89.8 | 87.7 | 88.7 | 87.8 | 89.7 | **87.7** | **89.4** |
| | 5 | 73.0 | 77.3 | 79.7 | 80.5 | 77.9 | 80.1 | **76.9** | **79.3** | 70.2 | 71.3 | 75.8 | 78.6 | 83.8 | 84.5 | 86.1 | 86.5 | **79.0** | **80.2** | 85.9 | 86.1 | 85.2 | 85.6 | 83.7 | 83.8 | 84.8 | 85.3 | **84.9** | **85.2** |
| | 6 | 74.6 | 76.7 | 76.0 | 80.1 | 72.2 | 78.6 | **74.3** | **78.5** | 61.6 | 68.8 | 74.0 | 75.8 | 84.1 | 83.4 | 86.7 | 86.0 | **76.6** | **78.5** | 85.4 | 85.1 | 86.8 | 85.4 | 85.3 | 83.5 | 85.1 | 84.4 | **85.7** | **84.6** |
| | 7 | 63.2 | 67.3 | 68.6 | 71.2 | 62.2 | 75.1 | **64.7** | **71.2** | 60.5 | 63.1 | 66.4 | 70.3 | 74.9 | 77.4 | 75.1 | 77.1 | **69.2** | **72.0** | 76.6 | 77.1 | 77.0 | 77.8 | 75.2 | 77.2 | 74.5 | 80.7 | **75.8** | **78.2** |
| | 8 | 66.5 | 67.2 | 68.3 | 71.2 | 60.6 | 74.7 | **65.1** | **71.0** | 55.4 | 60.9 | 66.4 | 67.5 | 76.7 | 76.9 | 77.3 | 76.9 | **69.0** | **70.5** | 78.4 | 77.3 | 79.4 | 77.5 | 77.7 | 76.9 | 77.6 | 80.9 | **78.3** | **78.1** |
| | 9 | 63.4 | 69.8 | 71.1 | 73.9 | 63.2 | 79.6 | **65.9** | **74.4** | 61.5 | 66.1 | 66.8 | 72.9 | 75.8 | 79.1 | 78.0 | 79.8 | **70.5** | **74.4** | 79.3 | 79.5 | 78.6 | 79.3 | 77.8 | 80.0 | 77.2 | 83.9 | **78.2** | **80.7** |
| | 10 | 67.4 | 70.0 | 70.7 | 74.1 | 59.2 | 77.0 | **65.8** | **73.7** | 55.6 | 64.7 | 65.7 | 69.4 | 78.1 | 78.6 | 80.4 | 79.8 | **70.0** | **73.1** | 80.6 | 79.1 | 81.2 | 79.0 | 80.5 | 79.7 | 80.0 | 83.9 | **80.6** | **80.4** |
| | 11 | 58.0 | 60.3 | 63.0 | 64.9 | 56.9 | 66.2 | **59.3** | **63.8** | 55.0 | 55.9 | 61.3 | 63.3 | 67.9 | 69.7 | 68.8 | 70.6 | **63.2** | **64.9** | 70.0 | 72.2 | 70.9 | 72.3 | 68.7 | 71.0 | 67.8 | 70.3 | **69.4** | **71.4** |
| | 12 | 60.5 | 59.6 | 63.1 | 64.6 | 56.0 | 65.5 | **59.9** | **63.3** | 51.8 | 54.5 | 61.5 | 59.9 | 69.5 | 68.2 | 70.7 | 69.3 | **63.4** | **63.0** | 72.1 | 72.1 | 73.6 | 71.9 | 71.7 | 70.6 | 70.6 | 69.6 | **72.0** | **71.1** |

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



**TABLE 12.** F1 score (F) and recognition accuracy (A) for each recognition task obtained usint Dataset-IG (GPS available).

| | | User1 | | User2 | | User3 | | Avg | | Pos1 | | Pos2 | | Pos3 | | Pos4 | | Avg | | Fold1 | | Fold2 | | Fold3 | | Fold4 | | Avg | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Scenario | F | R | F | R | F | R | F | R | F | R | F | R | F | R | F | R | F | R | F | R | F | R | F | R | F | R | F | R |
| A | 1 | 90.0 | 90.4 | 89.2 | 90.1 | 83.7 | 84.5 | 87.6 | 88.3 | 84.4 | 86.0 | 94.2 | 94.4 | 91.0 | 91.4 | 92.5 | 93.0 | 90.5 | 91.2 | 91.9 | 92.6 | 93.1 | 93.4 | 92.8 | 93.0 | 92.1 | 92.7 | 92.5 | 92.9 |
| | 2 | 76.1 | 83.2 | 77.6 | 81.9 | 74.9 | 77.0 | 76.2 | 80.7 | 72.8 | 77.7 | 82.6 | 88.4 | 73.6 | 82.2 | 86.6 | 88.9 | 78.9 | 84.3 | 82.4 | 86.0 | 82.5 | 88.0 | 81.6 | 86.6 | 80.5 | 85.5 | 81.7 | 86.5 |
| | 3 | 65.1 | 75.2 | 74.8 | 80.4 | 65.3 | 71.7 | 68.4 | 75.8 | 67.4 | 73.8 | 73.6 | 79.4 | 74.7 | 80.9 | 83.7 | 86.3 | 74.8 | 80.1 | 80.8 | 84.6 | 80.8 | 85.7 | 81.1 | 84.4 | 78.2 | 83.0 | 80.2 | 84.4 |
| | 4 | 69.4 | 75.0 | 75.6 | 80.1 | 63.7 | 70.1 | 69.6 | 75.1 | 69.8 | 72.3 | 78.8 | 81.3 | 75.9 | 79.3 | 85.8 | 86.2 | 77.6 | 79.8 | 81.7 | 83.6 | 84.1 | 85.8 | 83.6 | 84.0 | 81.5 | 83.0 | 82.7 | 84.1 |
| | 5 | 61.2 | 66.4 | 67.0 | 69.0 | 58.9 | 61.1 | 62.4 | 65.5 | 60.5 | 62.6 | 65.7 | 67.9 | 68.9 | 72.7 | 76.8 | 78.5 | 68.0 | 70.4 | 73.7 | 76.6 | 75.2 | 78.7 | 74.3 | 76.5 | 71.6 | 72.6 | 73.7 | 76.1 |
| | 6 | 65.6 | 66.2 | 68.4 | 69.1 | 60.7 | 62.3 | 64.9 | 65.9 | 64.5 | 61.6 | 72.9 | 72.4 | 71.4 | 71.4 | 79.4 | 78.0 | 72.1 | 70.9 | 75.0 | 74.8 | 78.8 | 78.8 | 77.7 | 76.4 | 74.2 | 71.2 | 76.4 | 75.3 |
| | 7 | 53.9 | 56.3 | 57.6 | 59.7 | 49.7 | 56.6 | 53.7 | 57.5 | 52.2 | 52.6 | 57.2 | 57.9 | 58.4 | 61.8 | 67.4 | 69.5 | 58.8 | 60.5 | 65.6 | 66.7 | 67.2 | 68.9 | 66.3 | 68.7 | 63.2 | 67.2 | 65.6 | 67.9 |
| | 8 | 57.8 | 55.5 | 61.2 | 59.9 | 50.3 | 56.3 | 56.4 | 57.2 | 57.6 | 53.1 | 62.3 | 59.0 | 63.2 | 61.5 | 71.5 | 69.7 | 63.7 | 60.8 | 68.2 | 65.8 | 71.1 | 68.9 | 70.2 | 68.6 | 68.3 | 67.8 | 69.5 | 67.7 |
| | 9 | 54.8 | 61.4 | 62.9 | 66.1 | 53.3 | 65.9 | 57.0 | 64.5 | 56.2 | 59.5 | 60.6 | 63.9 | 63.4 | 67.4 | 70.1 | 74.1 | 62.6 | 66.2 | 70.7 | 71.7 | 70.8 | 72.7 | 70.7 | 73.4 | 68.0 | 75.3 | 70.0 | 73.3 |
| | 10 | 62.1 | 62.3 | 65.3 | 66.2 | 53.5 | 64.1 | 60.3 | 64.2 | 61.8 | 59.7 | 65.2 | 64.1 | 68.4 | 68.2 | 76.0 | 75.3 | 67.8 | 66.8 | 72.8 | 70.8 | 75.4 | 73.2 | 74.2 | 72.9 | 73.1 | 75.6 | 73.9 | 73.1 |
| | 11 | 46.8 | 53.8 | 50.7 | 57.8 | 43.9 | 55.6 | 47.2 | 55.8 | 45.6 | 50.1 | 51.5 | 57.9 | 52.6 | 61.2 | 57.7 | 66.3 | 51.9 | 58.9 | 57.3 | 65.8 | 59.9 | 67.5 | 58.8 | 67.2 | 55.2 | 63.6 | 57.8 | 66.1 |
| | 12 | 51.9 | 53.9 | 54.1 | 58.3 | 44.6 | 53.8 | 50.2 | 55.3 | 52.0 | 51.1 | 55.5 | 57.1 | 56.8 | 60.6 | 63.5 | 66.9 | 57.0 | 58.9 | 60.5 | 64.9 | 63.8 | 67.0 | 62.8 | 66.6 | 59.8 | 63.6 | 61.7 | 65.5 |
| AG | 1 | 94.3 | 94.6 | 92.9 | 93.5 | 88.5 | 88.9 | 91.9 | 92.3 | 88.6 | 89.3 | 95.5 | 95.7 | 94.6 | 94.8 | 94.9 | 95.2 | 93.4 | 93.7 | 95.4 | 95.8 | 95.4 | 95.6 | 95.0 | 95.1 | 95.3 | 95.5 | 95.3 | 95.5 |
| | 2 | 82.1 | 88.3 | 80.1 | 85.7 | 81.4 | 83.5 | 81.2 | 85.8 | 74.9 | 81.3 | 85.1 | 90.9 | 81.5 | 88.0 | 88.5 | 90.8 | 82.5 | 87.8 | 86.1 | 90.0 | 85.5 | 90.7 | 84.1 | 89.4 | 83.9 | 88.8 | 84.9 | 89.7 |
| | 3 | 71.5 | 80.4 | 77.9 | 83.6 | 73.4 | 77.1 | 74.3 | 80.4 | 66.1 | 72.5 | 73.2 | 80.0 | 80.6 | 85.4 | 86.7 | 89.0 | 76.6 | 81.7 | 85.1 | 88.1 | 84.2 | 88.4 | 83.3 | 86.8 | 83.1 | 86.8 | 83.9 | 87.5 |
| | 4 | 72.2 | 78.4 | 78.0 | 84.0 | 67.6 | 74.3 | 72.6 | 78.9 | 57.5 | 70.5 | 69.0 | 73.7 | 81.0 | 84.0 | 87.7 | 88.7 | 73.8 | 79.2 | 84.8 | 87.1 | 86.5 | 88.5 | 85.7 | 86.8 | 85.2 | 86.6 | 85.5 | 87.2 |
| | 5 | 64.9 | 70.8 | 71.5 | 73.5 | 69.3 | 70.7 | 68.6 | 71.6 | 59.8 | 61.7 | 66.4 | 69.5 | 74.5 | 78.0 | 79.6 | 81.0 | 70.1 | 72.5 | 78.2 | 81.3 | 78.7 | 82.0 | 77.6 | 79.9 | 76.3 | 77.0 | 77.7 | 80.1 |
| | 6 | 67.3 | 69.2 | 71.0 | 73.3 | 66.3 | 69.8 | 68.2 | 70.8 | 54.0 | 59.7 | 69.3 | 69.5 | 75.6 | 76.1 | 82.5 | 81.4 | 70.4 | 71.7 | 79.0 | 79.9 | 81.5 | 81.8 | 80.4 | 79.8 | 78.5 | 75.9 | 79.9 | 79.4 |
| | 7 | 56.1 | 58.6 | 62.2 | 63.9 | 54.9 | 62.8 | 57.7 | 61.7 | 53.3 | 53.7 | 59.8 | 61.1 | 64.2 | 66.8 | 70.7 | 72.6 | 62.0 | 63.6 | 69.8 | 70.9 | 71.0 | 72.5 | 69.6 | 71.9 | 68.1 | 71.9 | 69.6 | 71.8 |
| | 8 | 60.1 | 58.1 | 62.8 | 63.0 | 54.3 | 62.0 | 59.1 | 61.0 | 50.2 | 52.4 | 61.1 | 58.8 | 67.6 | 66.3 | 73.1 | 71.8 | 63.0 | 62.3 | 72.6 | 70.6 | 74.3 | 72.2 | 73.3 | 72.0 | 72.1 | 72.0 | 73.1 | 71.7 |
| | 9 | 58.9 | 64.3 | 65.2 | 68.5 | 57.9 | 70.1 | 60.7 | 67.6 | 55.5 | 58.7 | 58.4 | 62.7 | 70.1 | 73.5 | 76.5 | 78.6 | 65.1 | 68.4 | 75.0 | 75.4 | 74.4 | 75.9 | 74.4 | 76.6 | 72.7 | 78.8 | 74.1 | 76.7 |
| | 10 | 62.2 | 62.9 | 67.1 | 69.2 | 56.6 | 69.9 | 61.9 | 67.3 | 52.6 | 58.1 | 59.2 | 59.0 | 72.3 | 72.0 | 78.5 | 78.2 | 65.6 | 66.8 | 77.2 | 75.1 | 78.2 | 76.1 | 77.6 | 76.6 | 76.3 | 78.9 | 77.3 | 76.7 |
| | 11 | 49.9 | 56.9 | 55.1 | 61.7 | 48.1 | 60.7 | 51.0 | 59.8 | 47.0 | 52.6 | 50.5 | 55.8 | 58.3 | 66.2 | 61.2 | 69.7 | 54.2 | 61.1 | 61.5 | 70.2 | 63.8 | 71.2 | 62.2 | 70.3 | 59.4 | 67.8 | 61.7 | 69.8 |
| | 12 | 52.1 | 54.6 | 56.7 | 61.5 | 48.3 | 58.9 | 52.4 | 58.3 | 44.5 | 50.2 | 54.8 | 57.0 | 61.0 | 64.7 | 64.7 | 69.0 | 56.3 | 60.2 | 64.8 | 69.7 | 67.2 | 70.7 | 65.8 | 69.9 | 63.1 | 67.2 | 65.2 | 69.3 |
| AGM | 1 | 95.5 | 95.8 | 93.7 | 94.2 | 91.0 | 91.2 | 93.4 | 93.7 | 89.8 | 90.4 | 96.5 | 96.6 | 95.1 | 95.3 | 95.4 | 95.7 | 94.2 | 94.5 | 95.7 | 96.0 | 95.8 | 96.0 | 95.3 | 95.5 | 95.9 | 96.1 | 95.7 | 95.9 |
| | 2 | 85.1 | 90.5 | 84.6 | 89.2 | 83.4 | 86.3 | 84.4 | 88.7 | 80.8 | 85.9 | 87.4 | 92.5 | 86.9 | 91.2 | 89.1 | 91.5 | 86.1 | 90.3 | 88.9 | 91.9 | 87.8 | 92.4 | 85.6 | 90.5 | 88.0 | 91.8 | 87.6 | 91.6 |
| | 3 | 74.6 | 82.6 | 81.1 | 85.9 | 75.1 | 79.0 | 76.9 | 82.5 | 69.7 | 76.4 | 73.5 | 79.8 | 85.0 | 88.6 | 87.2 | 89.7 | 78.9 | 83.6 | 87.3 | 90.0 | 86.2 | 90.1 | 84.8 | 88.1 | 86.3 | 89.7 | 86.1 | 89.5 |
| | 4 | 75.1 | 81.2 | 82.3 | 86.6 | 72.3 | 79.4 | 76.5 | 82.4 | 60.9 | 74.4 | 75.1 | 79.5 | 85.1 | 87.5 | 87.8 | 89.5 | 77.2 | 82.7 | 86.6 | 88.9 | 88.0 | 90.0 | 86.7 | 87.9 | 87.3 | 89.2 | 87.1 | 89.0 |
| | 5 | 71.7 | 76.1 | 79.6 | 81.4 | 74.2 | 75.1 | 75.2 | 77.5 | 67.6 | 69.0 | 74.3 | 76.8 | 82.0 | 84.3 | 84.6 | 85.5 | 77.1 | 78.9 | 83.6 | 85.5 | 84.0 | 86.3 | 81.9 | 83.7 | 83.2 | 84.1 | 83.2 | 84.9 |
| | 6 | 73.4 | 75.5 | 76.5 | 80.8 | 70.7 | 75.3 | 73.5 | 77.2 | 61.2 | 67.6 | 72.8 | 73.1 | 82.5 | 83.1 | 85.5 | 85.1 | 75.5 | 77.2 | 84.0 | 84.9 | 85.9 | 86.1 | 84.1 | 83.5 | 84.5 | 83.5 | 84.6 | 84.5 |
| | 7 | 61.4 | 63.4 | 68.5 | 69.7 | 60.2 | 69.2 | 63.4 | 67.4 | 60.1 | 60.3 | 62.4 | 62.5 | 73.2 | 74.6 | 73.7 | 75.5 | 67.3 | 68.2 | 75.5 | 75.6 | 75.5 | 76.2 | 73.9 | 75.7 | 73.2 | 77.7 | 74.5 | 76.3 |
| | 8 | 64.8 | 63.1 | 69.5 | 69.3 | 58.0 | 68.1 | 64.1 | 66.8 | 55.7 | 58.5 | 66.8 | 64.6 | 74.9 | 73.6 | 76.0 | 75.2 | 68.4 | 68.0 | 77.1 | 75.3 | 78.4 | 76.1 | 76.9 | 75.6 | 76.4 | 77.7 | 77.2 | 76.2 |
| | 9 | 62.8 | 67.3 | 68.8 | 71.2 | 60.1 | 73.4 | 63.9 | 70.6 | 59.5 | 62.2 | 61.6 | 65.2 | 74.9 | 77.3 | 77.1 | 79.5 | 68.3 | 71.0 | 77.4 | 77.6 | 76.5 | 77.6 | 75.9 | 78.1 | 74.9 | 81.5 | 76.2 | 78.7 |
| | 10 | 65.4 | 66.1 | 70.4 | 72.1 | 57.8 | 72.0 | 64.5 | 70.1 | 55.5 | 61.2 | 64.4 | 64.3 | 76.3 | 75.7 | 78.9 | 78.5 | 68.8 | 69.9 | 79.1 | 77.2 | 79.6 | 77.6 | 79.0 | 77.9 | 78.3 | 81.6 | 79.0 | 78.6 |
| | 11 | 56.3 | 60.9 | 62.3 | 67.2 | 53.7 | 64.8 | 57.4 | 64.3 | 54.4 | 57.8 | 56.4 | 58.5 | 66.6 | 72.0 | 66.1 | 72.3 | 60.9 | 65.1 | 67.8 | 73.6 | 68.8 | 73.7 | 67.2 | 73.4 | 65.2 | 71.7 | 67.3 | 73.1 |
| | 12 | 58.3 | 59.2 | 62.1 | 66.2 | 52.6 | 63.5 | 57.7 | 63.0 | 50.1 | 54.4 | 60.5 | 60.6 | 68.5 | 70.7 | 67.9 | 71.2 | 61.7 | 64.2 | 69.7 | 72.7 | 71.6 | 73.4 | 70.2 | 72.9 | 68.8 | 71.6 | 70.1 | 72.6 |
| P | 1 | 83.8 | 85.3 | 90.6 | 91.1 | 91.4 | 91.5 | 88.6 | 89.3 | 88.9 | 89.7 | 88.6 | 88.9 | 88.1 | 88.5 | 89.5 | 90.0 | 88.8 | 89.3 | 90.2 | 90.8 | 87.5 | 88.1 | 87.3 | 87.8 | 89.9 | 90.3 | 88.7 | 89.3 |
| | 2 | 82.1 | 83.2 | 86.2 | 88.6 | 88.2 | 89.8 | 85.5 | 87.2 | 85.0 | 87.1 | 85.1 | 87.1 | 84.2 | 86.3 | 87.3 | 88.2 | 85.4 | 87.2 | 86.7 | 88.4 | 84.4 | 86.1 | 84.2 | 85.7 | 86.5 | 88.4 | 85.5 | 87.1 |
| | 3 | 69.7 | 76.1 | 80.2 | 84.5 | 69.0 | 77.3 | 73.0 | 79.3 | 76.1 | 81.8 | 77.1 | 81.5 | 75.9 | 80.8 | 77.8 | 82.3 | 76.7 | 81.6 | 77.2 | 82.5 | 76.4 | 81.3 | 76.7 | 80.5 | 75.6 | 81.8 | 76.5 | 81.5 |
| | 4 | 57.1 | 74.8 | 65.6 | 83.2 | 57.8 | 74.2 | 60.2 | 77.4 | 65.5 | 80.2 | 64.9 | 79.4 | 63.8 | 78.6 | 66.6 | 80.3 | 65.2 | 79.6 | 66.8 | 80.4 | 64.1 | 79.3 | 63.8 | 78.8 | 64.8 | 79.5 | 64.9 | 79.5 |
| | 5 | 58.4 | 60.8 | 67.0 | 69.8 | 55.8 | 63.2 | 60.4 | 64.6 | 64.1 | 67.3 | 65.2 | 69.6 | 64.6 | 69.3 | 64.7 | 68.7 | 64.7 | 68.7 | 66.2 | 70.2 | 65.6 | 70.3 | 65.5 | 68.9 | 59.2 | 61.6 | 64.1 | 67.7 |
| | 6 | 49.7 | 59.2 | 57.2 | 68.5 | 48.1 | 59.7 | 51.6 | 62.5 | 57.5 | 65.7 | 56.8 | 67.4 | 56.6 | 67.1 | 57.4 | 66.8 | 57.1 | 66.8 | 59.4 | 68.1 | 56.7 | 68.2 | 57.3 | 67.2 | 52.7 | 59.4 | 56.6 | 65.7 |
| | 7 | 52.7 | 52.6 | 59.1 | 61.7 | 47.0 | 58.9 | 53.0 | 57.8 | 60.6 | 61.4 | 60.3 | 62.5 | 59.7 | 62.0 | 60.9 | 63.5 | 60.3 | 62.4 | 60.3 | 62.4 | 60.1 | 61.6 | 61.1 | 63.1 | 53.6 | 56.8 | 58.8 | 61.0 |
| | 8 | 46.3 | 51.4 | 51.8 | 60.4 | 42.0 | 55.8 | 46.7 | 55.8 | 55.2 | 59.8 | 54.0 | 60.3 | 53.6 | 59.9 | 55.1 | 61.5 | 54.5 | 60.4 | 55.1 | 60.1 | 53.3 | 59.5 | 54.6 | 61.3 | 48.9 | 54.6 | 53.0 | 58.9 |
| | 9 | 54.3 | 59.4 | 62.8 | 68.0 | 52.9 | 69.3 | 56.7 | 65.5 | 65.2 | 68.8 | 64.0 | 68.3 | 63.2 | 67.6 | 65.7 | 71.6 | 64.5 | 69.1 | 64.1 | 67.6 | 62.5 | 65.9 | 63.6 | 67.9 | 62.5 | 69.5 | 63.2 | 67.7 |
| | 10 | 46.5 | 58.1 | 53.7 | 66.7 | 45.9 | 66.3 | 48.7 | 63.7 | 58.3 | 67.2 | 55.9 | 66.1 | 55.6 | 65.4 | 58.5 | 69.6 | 57.1 | 67.1 | 57.4 | 65.3 | 54.5 | 63.9 | 55.9 | 66.2 | 55.5 | 67.2 | 55.8 | 65.7 |
| | 11 | 45.8 | 52.2 | 53.1 | 63.0 | 40.1 | 57.4 | 46.3 | 57.5 | 51.4 | 60.4 | 51.8 | 62.2 | 51.6 | 62.3 | 51.8 | 62.4 | 51.6 | 61.8 | 53.5 | 64.0 | 51.4 | 61.4 | 53.0 | 63.3 | 46.2 | 55.8 | 51.0 | 61.1 |
| | 12 | 40.8 | 50.9 | 47.6 | 61.7 | 36.3 | 53.9 | 41.6 | 55.5 | 47.9 | 58.8 | 47.3 | 60.1 | 47.2 | 60.1 | 48.1 | 60.4 | 47.6 | 59.8 | 50.1 | 61.9 | 46.6 | 59.3 | 48.4 | 61.7 | 42.9 | 53.5 | 47.0 | 59.1 |
| AP | 1 | 94.3 | 94.6 | 93.7 | 94.2 | 93.2 | 93.3 | 93.7 | 94.0 | 91.9 | 92.5 | 96.6 | 96.7 | 95.3 | 95.4 | 96.0 | 96.2 | 94.9 | 95.2 | 96.4 | 96.6 | 95.9 | 96.1 | 95.5 | 95.6 | 96.2 | 96.4 | 96.0 | 96.2 |
| | 2 | 88.9 | 91.5 | 89.6 | 92.0 | 91.3 | 92.4 | 89.9 | 92.0 | 87.8 | 90.4 | 92.1 | 94.9 | 90.6 | 92.9 | 93.1 | 94.5 | 90.9 | 93.2 | 92.4 | 94.3 | 91.6 | 94.1 | 91.4 | 93.6 | 92.1 | 94.3 | 91.9 | 94.1 |
| | 3 | 77.8 | 84.9 | 87.2 | 90.2 | 86.7 | 88.8 | 83.9 | 88.0 | 83.8 | 87.2 | 85.8 | 89.2 | 89.2 | 91.6 | 90.8 | 92.5 | 87.4 | 90.2 | 91.4 | 93.1 | 90.1 | 92.6 | 90.2 | 92.1 | 90.6 | 92.9 | 90.5 | 92.7 |
| | 4 | 80.8 | 85.0 | 83.2 | 89.8 | 83.8 | 88.7 | 82.6 | 87.8 | 84.6 | 86.4 | 86.7 | 89.7 | 90.2 | 91.4 | 91.9 | 92.5 | 88.3 | 90.0 | 90.6 | 92.5 | 91.3 | 92.6 | 91.0 | 91.8 | 91.5 | 92.8 | 91.1 | 92.4 |
| | 5 | 77.8 | 80.1 | 79.5 | 79.8 | 82.6 | 83.4 | 80.0 | 81.1 | 78.3 | 78.6 | 80.1 | 81.2 | 84.0 | 85.6 | 87.3 | 87.8 | 82.4 | 83.3 | 85.6 | 86.9 | 86.5 | 88.1 | 85.4 | 86.2 | 85.4 | 85.3 | 85.7 | 86.6 |
| | 6 | 77.7 | 78.2 | 76.6 | 78.4 | 79.4 | 81.9 | 77.9 | 79.5 | 79.6 | 77.8 | 83.2 | 83.5 | 85.6 | 85.1 | 89.0 | 88.0 | 84.3 | 83.6 | 85.7 | 85.9 | 88.0 | 88.1 | 86.9 | 86.1 | 86.8 | 84.7 | 86.9 | 86.2 |
| | 7 | 70.2 | 71.0 | 71.5 | 72.8 | 69.1 | 80.7 | 70.3 | 74.8 | 72.0 | 72.2 | 74.2 | 74.5 | 78.4 | 79.7 | 79.9 | 81.6 | 76.1 | 77.0 | 80.3 | 80.9 | 80.5 | 81.0 | 80.2 | 81.5 | 77.5 | 81.3 | 79.6 | 81.2 |
| | 8 | 71.9 | 69.7 | 71.1 | 72.4 | 68.4 | 79.4 | 70.5 | 73.9 | 74.8 | 71.8 | 78.0 | 77.1 | 80.3 | 78.9 | 82.4 | 81.8 | 78.9 | 77.4 | 81.0 | 79.9 | 82.9 | 81.2 | 82.1 | 81.1 | 80.4 | 81.4 | 81.6 | 80.9 |
| | 9 | 71.4 | 73.1 | 75.8 | 77.4 | 69.2 | 84.6 | 72.2 | 78.4 | 74.4 | 75.6 | 75.3 | 77.0 | 81.6 | 82.4 | 82.7 | 84.7 | 78.5 | 79.9 | 83.7 | 83.5 | 82.9 | 83.2 | 82.8 | 84.1 | 79.8 | 85.6 | 82.3 | 84.1 |
| | 10 | 71.7 | 70.5 | 73.6 | 76.2 | 68.1 | 83.1 | 71.1 | 76.6 | 77.2 | 75.4 | 79.5 | 79.6 | 84.1 | 82.7 | 83.4 | 83.6 | 81.0 | 80.3 | 84.1 | 82.6 | 85.1 | 83.1 | 84.4 | 83.6 | 82.7 | 85.5 | 84.1 | 83.7 |
| | 11 | 62.8 | 67.2 | 65.4 | 71.2 | 63.0 | 78.2 | 63.7 | 72.2 | 65.0 | 69.1 | 69.2 | 74.0 | 71.2 | 77.8 | 72.4 | 79.0 | 69.5 | 75.0 | 72.4 | 79.5 | 74.9 | 79.9 | 72.9 | 79.6 | 69.4 | 76.6 | 72.4 | 78.9 |
| | 12 | 65.8 | 66.7 | 65.1 | 70.7 | 62.3 | 75.7 | 64.4 | 71.0 | 68.5 | 69.4 | 71.8 | 74.8 | 73.9 | 77.3 | 74.7 | 78.6 | 72.2 | 75.0 | 74.0 | 78.4 | 77.0 | 79.3 | 75.5 | 79.1 | 73.0 | 76.6 | 74.9 | 78.4 |
| AGMP | 1 | 93.4 | 93.8 | 95.7 | 96.0 | 93.6 | 93.7 | 94.2 | 94.5 | 92.9 | 93.4 | 96.9 | 97.0 | 95.8 | 96.0 | 96.3 | 96.5 | 95.5 | 95.7 | 96.8 | 97.0 | 96.3 | 96.5 | 96.0 | 96.2 | 96.5 | 96.7 | 96.4 | 96.6 |
| | 2 | 89.6 | 92.4 | 90.5 | 93.0 | 91.0 | 92.5 | 90.4 | 92.6 | 87.9 | 90.9 | 92.2 | 95.1 | 91.7 | 94.3 | 92.6 | 94.4 | 91.1 | 93.7 | 92.8 | 94.7 | 92.1 | 94.8 | 91.5 | 94.0 | 92.5 | 94.9 | 92.2 | 94.6 |
| | 3 | 80.2 | 86.4 | 88.5 | 91.3 | 87.1 | 89.3 | 85.3 | 89.0 | 82.1 | 85.8 | 87.7 | 90.9 | 90.6 | 92.8 | 91.4 | 93.2 | 88.0 | 90.7 | 92.0 | 93.7 | 91.2 | 93.5 | 90.5 | 92.5 | 91.7 | 93.8 | 91.3 | 93.4 |
| | 4 | 79.6 | 84.5 | 84.9 | 91.5 | 83.0 | 88.2 | 82.5 | 88.0 | 84.5 | 86.1 | 86.4 | 89.8 | 91.3 | 92.6 | 92.0 | 93.0 | 88.6 | 90.4 | 90.9 | 93.0 | 91.7 | 93.2 | 91.4 | 92.5 | 92.3 | 93.6 | 91.6 | 93.1 |
| | 5 | 80.5 | 82.3 | 83.3 | 84.4 | 85.6 | 86.0 | 83.1 | 84.2 | 77.9 | 78.4 | 86.3 | 87.6 | 87.4 | 88.8 | 89.3 | 89.8 | 85.2 | 86.1 | 88.4 | 89.6 | 88.6 | 90.0 | 87.4 | 88.4 | 88.3 | 88.6 | 88.2 | 89.1 |
| | 6 | 79.2 | 80.1 | 81.4 | 85.1 | 80.6 | 84.2 | 80.4 | 83.1 | 81.0 | 79.2 | 83.6 | 84.5 | 88.7 | 88.7 | 89.7 | 89.2 | 85.7 | 85.4 | 88.2 | 88.9 | 89.9 | 90.1 | 88.4 | 88.0 | 89.5 | 88.5 | 89.0 | 88.9 |
| | 7 | 73.0 | 73.4 | 75.9 | 76.8 | 70.2 | 81.9 | 73.0 | 77.4 | 72.7 | 73.0 | 77.2 | 77.7 | 82.7 | 83.6 | 82.2 | 84.0 | 78.7 | 79.6 | 83.2 | 83.4 | 83.2 | 83.6 | 81.9 | 82.9 | 80.8 | 84.4 | 82.3 | 83.6 |
| | 8 | 73.6 | 71.6 | 75.3 | 76.6 | 68.6 | 80.9 | 72.5 | 76.4 | 75.8 | 73.5 | 80.0 | 79.4 | 84.1 | 83.0 | 83.5 | 83.4 | 80.9 | 79.8 | 83.6 | 82.6 | 84.8 | 83.3 | 83.8 | 82.9 | 83.1 | 84.5 | 83.8 | 83.3 |
| | 9 | 74.0 | 75.4 | 77.8 | 79.1 | 69.6 | 85.2 | 73.8 | 79.9 | 73.6 | 75.4 | 80.0 | 81.5 | 84.0 | 85.1 | 82.8 | 85.1 | 80.1 | 81.8 | 85.5 | 85.2 | 84.1 | 84.4 | 83.8 | 85.0 | 82.2 | 87.2 | 83.9 | 85.5 |
| | 10 | 72.7 | 71.9 | 76.8 | 79.4 | 67.6 | 83.5 | 72.3 | 78.3 | 76.8 | 75.4 | 78.7 | 79.1 | 86.1 | 85.0 | 84.9 | 85.3 | 81.6 | 81.2 | 85.6 | 84.4 | 86.1 | 84.4 | 85.6 | 84.8 | 84.6 | 87.4 | 85.5 | 85.2 |
| | 11 | 66.3 | 69.8 | 70.6 | 75.2 | 64.8 | 78.3 | 67.3 | 74.5 | 67.5 | 71.1 | 71.1 | 74.5 | 76.5 | 81.2 | 75.8 | 80.9 | 72.7 | 76.9 | 76.7 | 82.0 | 77.6 | 81.7 | 75.9 | 81.2 | 74.1 | 79.6 | 76.1 | 81.1 |
| | 12 | 68.1 | 68.4 | 70.2 | 74.6 | 65.1 | 77.9 | 67.8 | 73.6 | 70.7 | 70.8 | 75.6 | 77.3 | 78.5 | 80.7 | 77.4 | 80.0 | 75.5 | 77.2 | 77.9 | 81.1 | 79.9 | 81.3 | 78.5 | 81.1 | 76.7 | 79.2 | 78.2 | 80.7 |

frequency bands as well as the importance of magnetic field sensors are novel findings standing on their own irrespective of the classifier used, as they are the result of an information theoretical analysis.

There are a variety of ways to improve the recognition performance. Apart from using DT, we could use advanced classifiers, such as SVM and random forest. Post-filtering techniques, such as HMM and voting scheme, could be further employed to correct the prediction at individual frames. Some new features could be extracted from the sensor data, e.g. using deep learning, to further improve the recognition performance. In short, the improvement of the any proposed method could be identified easily by comparing with the baseline performance on the standard recognition tasks.

We perform feature computation and activity recognition with a sliding window of size 5.12 seconds. This window length is widely used in the related work and appears to be a good balance between decision time and accuracy. Ideally, the scientific community should standardize on a common window length, because the recognition performance varies significantly with the window length. However, if it is not possible to use a 5.12-second window size, other window lengths which are reported in the related work should ideally be used to enable comparison of methods. Researches using this dataset can always, based on their preference, establish their own baseline performance by targeting the recognition tasks defined in the paper. For instance, we think 60 seconds is also a good choice of window size, which is short enough for contextual awareness yet allows more complex GPS features.

In the SHL dataset the GPS information is not always available. Therefore, we evaluated different groups of modalities with two types of datasets: Dataset-E (the entire dataset) and Dataset-IG (the subset of Dataset-E where the GPS is available). In practice it may happen that the GPS is available sometimes and unavailable at other times. In this case it would be desirable to have two classifiers, one for when GPS is unavailable and one for when GPS is available, that can switch dynamically depending on the scenario. We would encourage the users to implement such a dynamic classifier and compare with the baseline results obtained with both Dataset-E and Dataset-IG.

The limited number of users might be a weak point of the SHL dataset. However, the variability in the sensor signal during transportation is primarily stemming from the motion of the vehicle as the movements of users within a vehicle are constrained (e.g. the movement of the bag containing the smartphone of two distinct users travelling in a bus would be quite similar). Therefore, when making the data collection protocol, we emphasized long travel distance and long duration recordings (over 7 months) at the expense of less users. We compensated this deficiency with rich sensor modalities (15 sensor modalities), multiple recording locations on body (4 locations), and high-quality annotations (28 context labels in total) [18], [19]. Meanwhile, we also realized the importance of having sufficient users and having a large geographical diversity in the dataset, so that the generality of the developed transportation mode recognition approaches can be verified with different people from different areas. Due to the limited time and funding, the data collection is confined mainly to the south of UK. Despite this, the SHL dataset is already one of the biggest datasets (in terms of duration, sensor modality and public availability) in the research community. We will continue improving the quality and size of the dataset in the future. By releasing this dataset and the tools to collect data, the scientific community can also contribute to expand it.

## VIII. CONCLUSIONS

In this paper we aim to advance the state-of-the-art research in transportation mode recognition by proposing standardized dataset, recognition tasks and evaluation criteria. We introduced a publicly available, large scale dataset (the Sussex-Huawei Locomotion dataset) for transportation mode recognition from multimodal smartphone sensors. The dataset consists of three users wearing four smartphones and conducting eight different transportation activities spanning seven months, leading to 2800 hours recording with 16 sensor modalities. The long duration, rich sensor modalities, the multiple users with various sensor placement, and the variety of transportation activities make the dataset a perfect candidate for establishing standard evaluation tasks. We recommended 12 reference scenarios which cover most recognition tasks identified in related work and defined three types of cross-validation measures including user-independent, sensor placement-independent and time-invariant evaluations. We suggested six relevant combinations of sensors to use based on energy considerations among accelerometer, gyroscope, magnetometer and GPS sensors. Taking advantage of the large amount of data, we computed a large number of statistical and frequency features in order to perform a systematic significance analysis based on the information theoretical criteria. We reported the reference performance on all the identified recognition scenarios with a machine-learning baseline. We provided guidelines on using the dataset and the defined recognition scenarios and evaluation criteria to generate reproducible and comparable results. We recommended researchers using the dataset to adhere to the tasks defined in this paper and refer to them with the name 'Task-Scenario-Crossvalidation-Modality'.

Through feature analysis we identified, for accelerometer, that important subband features mainly come from the frequency band between 0 and 10 Hz and compute both absolute energy and energy ratio; that important quantile features usually have a narrow interval between lower and upper quantiles; and that time-domain energy and time-domain kurtosis and frequency-domain energy are irrelevant features. We identified that, for gyroscope, important subband features mainly come from the frequency band between 0 and 30 Hz and compute both absolute energy and energy ratio; that important quantile features usually have a narrow

interval between lower and upper quantiles; and that time-domain energy and frequency-domain energy and frequency-domain kurtosis are irrelevant features. We identified, for magnetometer, that important subband features mainly come from the frequency band between 0 and 30 Hz and compute energy ratio only; that important quantile features usually have a narrow interval between lower and upper quantiles; and that time-domain energy and frequency-domain energy and the highest FFT value are irrelevant features.

The reference performance reported on the identified recognition scenarios demonstrates that advantages of using multiple modalities for transportation mode recognition. Particularly, the magnetometer modality is complementary to the accelerometer/gyroscope modality and combining the three can improve the recognition performance significantly over accelerometer and gyroscope. Similarly, combining GPS and accelerometer can also improve the recognition performance significantly over using accelerometer alone, and also over the combining of three inertial sensors.

We make the dataset and the baseline implementation publicly available to encourage a reproducible and fair comparison by the research community [19]. Future work would be to improve the recognition performance and to verify the generality of the SHL dataset by applying the classifier trained with the SHL dataset on other existing transportation mode recognition dataset.

## REFERENCES

[1] Y. Vaizman, K. Ellis, and G. Lanckriet, "Recognizing detailed human context in the wild from smartphones and smartwatches," IEEE Pervasive Computing, vol. 16, no. 4, pp. 62–74, Oct. 2017.

[2] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, "A survey of mobile phone sensing," IEEE Communications Magazine, vol. 48, no. 9, pp. 140–150, Sep. 2010.

[3] J. Wahlström, I. Skog, and P. Händel, "Smartphone-based vehicle telematics: A ten-year anniversary," IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 10, pp. 2802–2825, Apr. 2017.

[4] J. Engelbrecht, M. J. Booysen, G. J. van Rooyen, and F. J. Bruwer, "Survey of smartphone-based sensing in vehicles for intelligent transportation system applications," IET Intelligent Transport Systems, vol. 9, no. 10, pp. 924–935, Dec. 2015.

[5] C. Cottrill, F. Pereira, F. Zhao, I. Dias, H. Lim, M. Ben-Akiva, and P. Zegras, "Future mobility survey: Experience in developing a smartphone-based travel survey in singapore," Transportation Research Record: Journal of the Transportation Research Board, vol. 2354, pp. 59–67, Oct. 2013.

[6] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," IEEE Sensors Journal, vol. 15, no. 3, pp. 1321–1330, Dec. 2015.

[7] J. Froehlich, T. Dillahunt, P. Klasnja, J. Mankoff, S. Consolvo, B. Harrison, and J. A. Landay, "Ubigreen: Investigating a mobile tool for tracking and supporting green transportation habits," in Proc. SIGCHI Conference on Human Factors in Computing Systems, Boston, USA, 2009, pp. 1043–1052.

[8] W. Brazil and B. Caulfield, "Does green make a difference: The potential role of smartphone technology in transport behaviour," Transportation Research Part C: Emerging Technologies, vol. 37, pp. 93–101, Dec. 2013.

[9] E. Agapie, G. Chen, D. Houston, E. Howard, J. Kim, M. Y. Mun, A. Mondschein, S. Reddy, R. Rosario, J. Ryder et al., "Seeing our signals: Combining location traces and web-based models for personal discovery," in Proc. ACM Workshop on Mobile Computing Systems and Applications, Napa Valley, USA, 2008, pp. 6–10.

[10] C. Dobre and F. Xhafa, "Intelligent services for big data science," Future Generation Computer Systems, vol. 37, pp. 267–281, Jul. 2014.

[11] N. A. Streitz, "From human–computer interaction to human–environment interaction: Ambient intelligence and the disappearing computer," in Proc. Universal Access in Ambient Intelligence Environments, Königswinter, Germany, 2007, pp. 3–13.

[12] D. A. Johnson and M. M. Trivedi, "Driving style recognition using a smartphone as a sensor platform," in Proc. IEEE Conference on Intelligent Transportation Systems, Washington, USA, 2011, pp. 1609–1615.

[13] G. Castignani, T. Derrmann, R. Frank, and T. Engel, "Driver behavior profiling using smartphones: A low-cost platform for driver monitoring," IEEE Intelligent Transportation Systems Magazine, vol. 7, no. 1, pp. 91–102, Jan. 2015.

[14] J. Biancat, C. Brighenti, and A. Brighenti, "Review of transportation mode detection techniques," EAI Endorsed Transactions on Ambient Systems, vol. 1, no. 4, pp. 1–10, Jan. 2014.

[15] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes challenge: A retrospective," International Journal of Computer Vision, vol. 111, no. 1, pp. 98–136, Jan. 2015.

[16] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and F. Li, "Imagenet large scale visual recognition challenge," International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015.

[17] H. Christensen, J. Barker, N. Ma, and P. D. Green, "The CHiME corpus: A resource and a challenge for computational hearing in multisource environments," in Proc. InterSpeech Conference, Makuhari, Japan, 2010, pp. 1918–1921.

[18] H. Gjoreski, M. Ciliberto, L. Wang, F. J. O. Morales, S. Mekki, S. Valentin, and D. Roggen, "The University of Sussex-Huawei locomotion and transportation dataset for multimodal analytics with mobile devices," IEEE Access, vol. 6, pp. 42 592–42 604, 2018.

[19] SHL Dataset, http://www.shl-dataset.org, accessed Dec. 2017.

[20] S. H. Fang, H. H. Liao, Y. X. Fei, K. H. Chen, J. W. Huang, Y. D. Lu, and Y. Tsao, "Transportation modes classification using sensors on smartphones," Sensors, vol. 16, no. 8, pp. 1324–1339, Aug. 2016.

[21] S. H. Fang, Y. X. Fei, Z. Xu, and Y. Tsao, "Learning transportation modes from smartphone sensors based on deep neural network," IEEE Sensors Journal, vol. 17, no. 18, pp. 6111–6118, Aug. 2017.

[22] M. C. Yu, T. Yu, S. C. Wang, C. J. Lin, and E. Y. Chang, "Big data small footprint: The design of a low-power classifier for detecting transportation modes," in Proc. Very Large Data Base Endowment, Hangzhou, China, 2014, pp. 1429–1440.

[23] S. Wang, C. Chen, and J. Ma, "Accelerometer based transportation mode recognition on mobile phones," in Proc. Asia-Pacific Conference on Wearable Computing Systems, Shenzhen, China, 2010, pp. 44–46.

[24] A. Jahangiri and H. A. Rakha, "Applying machine learning techniques to transportation mode recognition using mobile phone sensor data," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 5, pp. 2406–2417, Mar. 2015.

[25] S. Hemminki, P. Nurmi, and S. Tarkoma, "Accelerometer-based transportation mode detection on smartphones," in Proc. ACM Conference on Embedded Networked Sensor Systems, Roma, Italy, 2013, pp. 1–14.

[26] X. Su, H. Caceres, H. Tong, and Q. He, "Online travel mode identification using smartphones with battery saving considerations," IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 10, pp. 2921–2934, Mar. 2016.

[27] P. Siirtola and J. Röning, "Recognizing human activities user-independently on smartphones based on accelerometer data," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 1, no. 5, pp. 38–45, Nov. 2012.

[28] Z. Zhang and S. Poslad, "A new post correction algorithm (pocoa) for improved transportation mode recognition," in Proc. IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 2013, pp. 1512–1518.

[29] M. A. Shafique and E. Hato, "Use of acceleration data for transportation mode prediction," Transportation, vol. 42, no. 1, pp. 163–188, Jan. 2015.

[30] D. Shin, D. Aliaga, B. Tunçer, S. M. Arisona, S. Kim, D. Zünd, and G. Schmitt, "Urban sensing: Using smartphones for transportation mode classification," Computers, Environment and Urban Systems, vol. 53, pp. 76–86, Sep. 2015.

[31] B. Nham, K. Siangliulue, and S. Yeung, "Predicting mode of transport from iphone accelerometer data," Tech. Rep., 2008.

[32] E. Hedemalm, "Online transportation mode recognition and an application to promote greener transportation," *Master Thesis*, Luleå University of Technology, Luleå, Sweden, 2017.

[33] J. Yang, "Toward physical activity diary: Motion recognition using simple acceleration features with mobile phones," in Proc. International

Workshop on Interactive Multimedia for Consumer Electronics, Beijing, China, 2009, pp. 1–10.
[34] T. Nick, E. Coersmeier, J. Geldmacher, and J. Goetze, "Classifying means of transportation using mobile sensor data," in Proc. Internetional Joint Conference on Neural Networks, Barcelona, Spain, 2010, pp. 1–6.
[35] T. Sonderen, "Detection of transportation mode solely using smartphones," in Proc. Twente Student Conference on IT, Enschede, Netherlands, 2016, pp. 1–7.
[36] V. Manzoni, D. Maniloff, K. Kloeckl, and C. Ratti, "Transportation mode identification and real-time CO2 emission estimation using smartphones," *SENSEable City Lab, Massachusetts Institute of Technology*, Massachusetts, USA, Tech. Rep., 2010.
[37] K. Sankaran, M. Zhu, X. F. Guo, A. L. Ananda, M. C. Chan, and L.-S. Peh, "Using mobile phone barometer for low-power transportation context detection," in Proc. ACM Conference on Embedded Networked Sensor Systems, Memphis, USA, 2014, pp. 191–205.
[38] Y. Zheng, Y. Chen, Q. Li, X. Xie, and W. Y. Ma, "Understanding transportation modes based on GPS data for web applications," ACM Transactions on the Web, vol. 4, no. 1, pp. 1–36, Jan. 2010.
[39] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W. Y. Ma, "Understanding mobility based on GPS data," in Proc. International Conference on Ubiquitous Computing, Seoul, Korea, 2008, pp. 312–321.
[40] Z. Xiao, Y. Wang, K. Fu, and F. Wu, "Identifying different transportation modes from trajectory data using tree-based ensemble classifiers," ISPRS International Journal of Geo-Information, vol. 6, no. 2, pp. 57–79, 2017.
[41] Y. Endo, H. Toda, K. Nishida, and J. Ikedo, "Classifying spatial trajectories using representation learning," International Journal of Data Science and Analytics, vol. 2, no. 3-4, pp. 107–117, Dec. 2016.
[42] G. Xiao, Z. Juan, and C. Zhang, "Travel mode detection based on GPS track data and bayesian networks," Computers, Environment and Urban Systems, vol. 54, pp. 14–22, Nov. 2015.
[43] L. Stenneth, O. Wolfson, P. S. Yu, and B. Xu, "Transportation mode detection using mobile phones and GIS information," in Proc. ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, Illinois, 2011, pp. 54–63.
[44] I. Semanjski, S. Gautama, R. Ahas, and F. Witlox, "Spatial context mining approach for transport mode recognition from mobile sensed big data," Computers, Environment and Urban Systems, vol. 66, pp. 38–52, Nov. 2017.
[45] A. Bolbol, T. Cheng, I. Tsapakis, and J. Haworth, "Inferring hybrid transportation modes from sparse GPS data using a moving window svm classification," Computers, Environment and Urban Systems, vol. 36, no. 6, pp. 526–537, Nov. 2012.
[46] H. Gong, C. Chen, E. Bialostozky, and C. T. Lawson, "A GPS/GIS method for travel mode detection in new york city," Computers, Environment and Urban Systems, vol. 36, no. 2, pp. 131–139, 2012.
[47] P. A. Gonzalez, J. S. Weinstein, S. J. Barbeau, M. A. Labrador, P. L. Winters, N. L. Georggi, and R. Perez, "Automating mode detection for travel behaviour analysis by using global positioning systems-enabled mobile phones and neural networks," IET Intelligent Transport Systems, vol. 4, no. 1, pp. 37–49, Mar. 2010.
[48] T. Feng and H. J. Timmermans, "Transportation mode recognition using GPS and accelerometer data," Transportation Research Part C: Emerging Technologies, vol. 37, pp. 118–130, Dec. 2013.
[49] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Using mobile phones to determine transportation modes," ACM Transactions on Sensor Networks, vol. 6, no. 2, pp. 1–27, Feb. 2010.
[50] P. Widhalm, P. Nitsche, and N. Brändie, "Transport mode detection with realistic smartphone sensor data," in Proc. International Conference on Pattern Recognition, Tsukuba, Japan, 2012, pp. 573–576.
[51] H. Lu, J. Yang, Z. Liu, N. D. Lane, T. Choudhury, and A. T. Campbell, "The jigsaw continuous sensing engine for mobile phone applications," in Proc. ACM Conference on Embedded Networked Sensor Systems, Zurich, Switzerland, 2010, pp. 71–84.
[52] P. Nitsche, P. Widhalm, S. Breuss, and P. Maurer, "A strategy on how to utilize smartphones for automatically reconstructing trips in travel surveys," Procedia-Social and Behavioral Sciences, vol. 48, pp. 1033–1046, Dec. 2012.
[53] H. Xia, Y. Qiao, J. Jian, and Y. Chang, "Using smart phone sensors to detect transportation modes," Sensors, vol. 14, no. 11, pp. 20 843–20 865, Nov. 2014.
[54] S. L. Lau and K. David, "Movement recognition using the accelerometer in smartphones," in Proc. Future Network and Mobile Summit, Florence, Italy, 2010, pp. 1–9.
[55] US Transportation Dataset, http://cs.unibo.it/projects/us-tm2017/index.html, accessed Nov. 2017.
[56] Y. Zheng, X. Xie, and W. Y. Ma, "Geolife: A collaborative social networking service among user, location and trajectory." IEEE Data Engineering Bulletin, vol. 33, no. 2, pp. 32–39, Jun. 2010.
[57] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," ACM Computing Surveys, vol. 46, no. 3, pp. 1–33, 2014.
[58] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1226–1238, Jun. 2005.
[59] N. Kwak and C.-H. Choi, "Input feature selection by mutual information based on parzen window," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 12, pp. 1667–1671, Dec. 2002.
[60] P. A. Estévez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," IEEE Transactions on Neural Networks, vol. 20, no. 2, pp. 189–201, Jan. 2009.

LIN WANG received the B.S. degree in electronic engineering from Tianjin University, China, in 2003; and the Ph.D degree in signal processing from Dalian University of Technology, China, in 2010. From 2011 to 2013, he has been an Alexander von Humboldt Fellow in University of Oldenburg, Germany. From 2014 to 2017, he has been a postdoctoral researcher in Queen Mary University of London, UK. From 2017 to 2018, he has been a postdoctoral researcher in the University of Sussex, UK. Since September 2018, he has been a Lecturer in Queen Mary University of London. His research interests include video and audio compression, blind source separation, 3D audio, and machine learning (https://sites.google.com/site/linwangsig/).

HRISTIJAN GJORESKI received the M.Sc. and Ph.D. degree in information and communication technologies from the Jozef Stefan Postgraduate School, Ljubljana, Slovenia, in 2011 and 2015. From 2010 to 2015, he was an Assistant Researcher at the Department of Intelligent Systems at the Jozef Stefan Institute in Ljubljana, Slovenia. Since 2016, he has been a Postdoctoral Research Fellow at the Sensor Technology Research Center, at the University of Sussex, United Kingdom. His research interests include Artificial Intelligence, Machine Learning, Wearable Computing, Time-series analysis. Dr. Gjoreski was a recipient of the Best Young Scientist for 2016, award given by the President of Macedonia. Additionally, he was part of the team that won the International EvAAL Activity Recognition Challenge in 2013.

MATHIAS CILIBERTO received his M.Sc. in Computer Science from the University of Milan, Milan, Italy, in 2015. In 2016, he joined the Wearable Technologies Lab as Ph.D. student under the supervision of Dr. Daniel Roggen, within the Sensor Technology Research Center, at the University of Sussex, United Kingdom. His research focuses on Wearable Technologies in sports, and his interests include Machine Learning, as well as Artificial Intelligence and Activity Recognition using Wearable Computing.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2890793, IEEE Access



SAMI MEKKI received the engineering diploma in Wireless Networks from SUPCOM, Tunis, Tunisia in 2004, the M.Sc degree in signal and digital communication from UNSA (university of Nice Sophia-Antipolis) in 2005 and the Ph.D degree in electrical engineering from Telecom ParisTech in 2009. He worked in different companies as well as for CNRS (Centre national de la recherche scientifique) and the French Atomic Energy Commission. He is currently a senior researcher at the Mathematical and Algorithmic Sciences Lab, PRC, Huawei Technologies France. His research interests include wireless communication, channel estimation, multiuser detection and sensor data fusion for user localizatoin.

STEFAN VALENTIN (S'07, M'09) graduated in EE from the Technical University of Berlin, Germany in 2004 and received his Ph.D. in CS with summa cum laude from the University of Paderborn, Germany in 2010. In the same year, he joined Bell Labs, Stuttgart, Germany as a Member of Technical Staff, where he worked on wireless resource allocation algorithms for 4G and 5G. From 2015 to Sep. 2018 he was with Huawei's Mathematical and Algorithmic Sciences Lab in Paris as Principal Researcher and team leader. Since Oct. 2018, he is full Professor at the Department of Computer Science, Darmstadt University of Applied Sciences in Germany. Stefan's research interests are wireless resource allocation and load balancing for 5G and beyond. His methodological interest reaches from mathematical optimization via Bayesian statistics to machine learning. In these fields, he received two best paper awards, various awards from industry, the Klaus Tschira Award in 2011, and IEEE ComSoc's Fred W. Ellersick Prize in 2015.

DANIEL ROGGEN (M'04) is Associate Professor in Sensor Technologies at the University of Sussex, where he leads the Wearable Technologies Lab and directs the Sensor Technology Research Centre. His research focuses on wearable and mobile computing, activity and context recognition, and intelligent embedded systems. He has established a number of recognized datasets for human activity recognition from wearable sensors, in particular the OPPORTUNITY dataset. He is member Task Force on Intelligent Cyber-Physical Systems of the of the IEEE Computational Intelligence Society. He received his Masters degree (2001) and PhD (2005) from the Ecole Polytechnique Federale de Lausanne, Switzerland.